

COMPOUNDS AND METHODS FOR TREATMENT  
AND DIAGNOSIS OF CHLAMYDIAL INFECTION

5

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates generally to the detection and treatment of Chlamydial infection. In particular, the invention is related to polypeptides comprising a  
10 *Chlamydia* antigen and the use of such polypeptides for the serodiagnosis and treatment of Chlamydial infection.

Description of Related Art

Chlamydiae are intracellular bacterial pathogens that are responsible for a  
15 wide variety of important human and animal infections. *Chlamydia trachomatis* is one of the most common causes of sexually transmitted diseases and can lead to pelvic inflammatory disease (PID), resulting in tubal obstruction and infertility. *Chlamydia trachomatis* may also play a role in male infertility. In 1990, the cost of treating PID in the US was estimated to be \$4 billion. Trachoma, due to ocular infection with *Chlamydia*  
20 *trachomatis*, is the leading cause of preventable blindness worldwide. *Chlamydia pneumonia* is a major cause of acute respiratory tract infections in humans and is also believed to play a role in the pathogenesis of atherosclerosis and, in particular, coronary heart disease. Individuals with a high titer of antibodies to *Chlamydia pneumonia* have been shown to be at least twice as likely to suffer from coronary heart disease as  
25 seronegative individuals. Chlamydial infections thus constitute a significant health problem both in the US and worldwide.

Chlamydial infection is often asymptomatic. For example, by the time a woman seeks medical attention for PID, irreversible damage may have already occurred resulting in infertility. There thus remains a need in the art for improved vaccines and  
30 pharmaceutical compositions for the prevention and treatment of *Chlamydia* infections. The present invention fulfills this need and further provides other related advantages.

## BRIEF SUMMARY OF THE INVENTION

The present invention provides compositions and methods for the diagnosis and therapy of *Chlamydia* infection. In one aspect, the present invention provides polypeptides comprising an immunogenic portion of a *Chlamydia* antigen, or a variant of such an antigen. Certain portions and other variants are immunogenic, such that the ability of the variant to react with antigen-specific antisera is not substantially diminished. Within certain embodiments, the polypeptide comprises an amino acid sequence encoded by a polynucleotide sequence selected from the group consisting of (a) a sequence of SEQ ID NO: 1-48, 114-121, 125-138, 141-157; (b) the complements of said sequences; and (c) sequences that hybridize to a sequence of (a) or (b) under moderately stringent conditions. In specific embodiments, the polypeptides of the present invention comprise at least a portion of a *Chlamydial* protein that includes an amino acid sequence selected from the group consisting of sequences recited in SEQ ID NO:122-124 and 139-140 and variants thereof.

The present invention further provides polynucleotides that encode a polypeptide as described above, or a portion thereof (such as a portion encoding at least 15 amino acid residues of a *Chlamydial* protein), expression vectors comprising such polynucleotides and host cells transformed or transfected with such expression vectors.

In a related aspect, polynucleotide sequences encoding the above polypeptides, recombinant expression vectors comprising one or more of these polynucleotide sequences and host cells transformed or transfected with such expression vectors are also provided.

In another aspect, the present invention provides fusion proteins comprising an inventive polypeptide, or, alternatively, an inventive polypeptide and a known *Chlamydia* antigen, as well as polynucleotides encoding such fusion proteins, in combination with a physiologically acceptable carrier or immunostimulant for use as pharmaceutical compositions and vaccines thereof.

The present invention further provides pharmaceutical compositions that comprise: (a) an antibody, both polyclonal and monoclonal, or antigen-binding fragment thereof that specifically binds to a *Chlamydial* protein; and (b) a physiologically acceptable

that comprise one or more *Chlamydia* polypeptides disclosed herein, for example, a polypeptide of SEQ ID NO: 95-109, 122-124 and 139-140, or a polynucleotide molecule encoding such a polypeptide, such as a polynucleotide sequence of SEQ ID NO: 1-48, 80-94, 114-121 125-138, and 141-157, and a physiologically acceptable carrier. The invention  
5 also provides compositions for prophylactic and therapeutic purposes comprising one or more of the disclosed polynucleotides and/or polypeptides and an immunostimulant, e.g., an adjuvant.

In yet another aspect, methods are provided for stimulating an immune response in a patient, e.g., for inducing protective immunity in a patient, comprising  
10 administering to a patient an effective amount of one or more of the above pharmaceutical compositions or vaccines.

In yet a further aspect, methods for the treatment of *Chlamydia* infection in a patient are provided, the methods comprising obtaining peripheral blood mononuclear cells (PBMC) from the patient, incubating the PBMC with a polypeptide of the present invention  
15 (or a polynucleotide that encodes such a polypeptide) to provide incubated T cells and administering the incubated T cells to the patient. The present invention additionally provides methods for the treatment of *Chlamydia* infection that comprise incubating antigen presenting cells with a polypeptide of the present invention (or a polynucleotide that encodes such a polypeptide) to provide incubated antigen presenting cells and  
20 administering the incubated antigen presenting cells to the patient. Proliferated cells may, but need not, be cloned prior to administration to the patient. In certain embodiments, the antigen presenting cells are selected from the group consisting of dendritic cells, macrophages, monocytes, B-cells, and fibroblasts. Compositions for the treatment of *Chlamydia* infection comprising T cells or antigen presenting cells that have been  
25 incubated with a polypeptide or polynucleotide of the present invention are also provided. Within related aspects, vaccines are provided that comprise: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) an immunostimulant.

The present invention further provides, within other aspects, methods for removing *Chlamydial*-infected cells from a biological sample, comprising contacting a  
30 biological sample with T cells that specifically react with a *Chlamydial* protein, wherein the step of contacting is performed under conditions and for a time sufficient to permit the

removal of cells expressing the protein from the sample.

Within related aspects, methods are provided for inhibiting the development of *Chlamydial* infection in a patient, comprising administering to a patient a biological sample treated as described above. In further aspects of the subject invention, methods and  
5 diagnostic kits are provided for detecting *Chlamydia* infection in a patient. In one embodiment, the method comprises: (a) contacting a biological sample with at least one of the polypeptides or fusion proteins disclosed herein; and (b) detecting in the sample the presence of binding agents that bind to the polypeptide or fusion protein, thereby detecting *Chlamydia* infection in the biological sample. Suitable biological samples include whole  
10 blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine. In one embodiment, the diagnostic kits comprise one or more of the polypeptides or fusion proteins disclosed herein in combination with a detection reagent. In yet another embodiment, the diagnostic kits comprise either a monoclonal antibody or a polyclonal antibody that binds with a polypeptide of the present invention.

15 The present invention also provides methods for detecting *Chlamydia* infection comprising: (a) obtaining a biological sample from a patient; (b) contacting the sample with at least two oligonucleotide primers in a polymerase chain reaction, at least one of the oligonucleotide primers being specific for a polynucleotide sequence disclosed herein; and (c) detecting in the sample a polynucleotide sequence that amplifies in the  
20 presence of the oligonucleotide primers. In one embodiment, the oligonucleotide primer comprises at least about 10 contiguous nucleotides of a polynucleotide sequence peptide disclosed herein, or of a sequence that hybridizes thereto.

In a further aspect, the present invention provides a method for detecting *Chlamydia* infection in a patient comprising: (a) obtaining a biological sample from the  
25 patient; (b) contacting the sample with an oligonucleotide probe specific for a polynucleotide sequence disclosed herein; and (c) detecting in the sample a polynucleotide sequence that hybridizes to the oligonucleotide probe. In one embodiment, the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a polynucleotide sequence disclosed herein, or a sequence that hybridizes thereto.



These and other aspects of the present invention will become apparent upon reference to the following detailed description. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

## 5 SEQUENCE IDENTIFIERS

SEQ ID NO:1 sets forth a DNA sequence identified for clone E4-A2-39 (CT10 positive) that is 1311 bp and contains the entire ORF for CT460 (SWIB) and a partial ORF for CT461 (yaeI).

10 SEQ ID NO:2 sets forth a DNA sequence for clone E2-B10-52 (CT10 positive) that has a 1516 bp insert that contains partial ORFs for genes CT827 (nrdA-ribonucleoside reductase large chain) and CT828 (nrdB-ribonucleoside reductase small chain). These genes were not identified in a Ct L2 library screening.

SEQ ID NO:3 sets forth a DNA sequence for clone E1-B1-80 (CT10 positive) (2397bp) that contains partial ORFs for several genes, CT812 (pmpD), CT015 (phoH ATPase), CT016 (hypothetical protein) and pGp1-D (C. trachomatis plasmid gene). 15

SEQ ID NO:4 sets forth a DNA sequence for clone E4-F9-4 (CT10, CL8, CT1, CT5, CT13, and CHH037 positive) that contains a 1094 bp insert that has a partial ORF for the gene CT316 (L7/L12 ribosomal protein) as well as a partial ORF for gene CT315 (RNA polymerase beta).

20 SEQ ID NO:5 sets forth a DNA sequence for clone E2-H6-40 (CT3 positive) that has a 2129 bp insert that contains the entire ORF for the gene CT288 and very small fragments of genes CT287 and CT289. Genes in this clone have not been identified in screening with a Ct L2 library.

SEQ ID NO:6 sets forth a DNA sequence for clone E5-D4-2 (CT3, CT10, CT1, CT5, CT12, and CHH037 positive) that has a 1828 bp insert that contains a partial ORF for gene CT378 (pgi), complete ORF for gene CT377 (ltuA) and a complete ORF for the gene CT376 (malate dehydrogenase). In addition, the patient lines CT10, CT1, CT5, CT12, and CHH037 also identified this clone. 25

SEQ ID NO:7 sets forth a DNA sequence for clone E6-C1-31 (CT3 positive) 30 that has a 861 bp insert that contains a partial ORF for gene CT858.

SEQ ID NO:8 sets forth a DNA sequence for clone E9-E11-76 (CT3 positive) that contains a 763 bp insert that is an amino terminal region of the gene for CT798 (Glycogen synthase). This gene was not identified in a previous screening with a Ct L2 library.

5           SEQ ID NO:9 sets forth a DNA sequence for clone E2-A9-26 (CT1-positive) that contains part of the gene for ORF-3 which is found on the plasmid in *Chlamydia trachomatis*.

          SEQ ID NO:10 sets forth a DNA sequence for clone E2-G8-94 (CT1-positive) that has the carboxy terminal end of Lpda gene as well as a partial ORF for  
10 CT556.

          SEQ ID NO: 11 sets forth a DNA sequence for clone E1-H1-14 (CT1 positive) that has a 1474 bp insert that contains the amino terminal part of an Lpda ORF on the complementary strand.

          SEQ ID NO: 12 sets forth a DNA sequence for clone E1-A5-53 (CT1  
15 positive) that contains a 2017 bp insert that has an amino terminal portion of the ORF for dnaK gene on the complementary strand, a partial ORF for the grpE gene (CT395) and a partial ORF for CT166 .

          SEQ ID NO: 13 sets forth a DNA sequence for clone E3-A1-50 (positive on CT1 line) that is 1199 bp and contains a carboxy terminal portion of the ORF for CT622.

20           SEQ ID NO: 14 sets forth a DNA sequence for clone E3-E2-22 that has 877 bp, containing a complete ORF for CT610 on the complementary strand, and was positive on both CT3 and CT10 lines.

          SEQ ID NO: 15 sets forth the DNA sequence for clone E5-E2-10 (CT10 positive) which is 427 bp and contains a partial ORF for the major outer membrane protein omp1. SEQ ID NO: 16 sets forth the DNA sequence for clone E2-D5-89 (516bp) which is a  
25 CT10 positive clone that contains a partial ORF for pmpD gene (CT812).

          SEQ ID NO: 17 sets forth the DNA sequence for clone E4-G9-75 (CT10 positive) which is 723 bp and contains a partial ORF for the amino terminal region of the pmpH gene (CT872).

SEQ ID NO: 18 sets forth the DNA sequence for clone E3-F2-37 (CT10, CT3, CT11, and CT13 positive-1377bp insert) which contains a partial ORF for the tRNA-Trp (CT322) gene and a complete ORF for the gene secE (CT321).

SEQ ID NO: 19 sets forth the DNA sequence for clone E5-A11-8 (CT10 positive-1736 bp) which contains the complete ORF for groES (CT111) and a majority of the ORF for groEL (CT110).

SEQ ID NO: 20 sets forth the DNA sequence for clone E7-H11-61 (CT3 positive-1135 bp) which has partial inserts for fliA (CT061), tyrS (CT062), TSA (CT603) and a hypothetical protein (CT602).

10 SEQ ID NO: 21 sets forth a DNA sequence for clone E6-C8-95 which contains a 731 bp insert that was identified using the donor lines CT3, CT1, and CT12 line. This insert has a carboxy terminal half for the gene for the 60 kDa ORF.

SEQ ID NO: 22 sets forth the DNA sequence for clone E4-D2-79 (CT3 positive) which contains a 1181 bp insert that is a partial ORF for nrdA gene. The ORF for 15 this gene was also identified from clone E2-B10-52 (CT10 positive).

SEQ ID NO: 23 sets forth the DNA sequence for clone E1-F9-79 (167 bp; CT11 positive) which contains a partial ORF for the gene CT133 on the complementary strand. CT133 is a predicted rRNA methylase.

SEQ ID NO: 24 sets forth the DNA sequence for clone E2-G12-52 (1265 20 bp; CT11 positive) which contains a partial ORF for clpB, a protease ATPase.

SEQ ID NO: 25 sets forth the DNA sequence for clone E4-H3-56 (463 bp insert; CT1 positive) which contains a partial ORF for the TSA gene (CT603) on the complementary strand.

SEQ ID NO: 26 sets forth the DNA sequence for clone E5-E9-3 (CT1 25 positive) that contains a 636 bp insert partially encoding the ORF for dnaK like gene. Part of this sequence was also identified in clone E1-A5-53.

SEQ ID NO: 27 sets forth the full-length serovar E DNA sequence of CT875.

SEQ ID NO: 28 sets forth the full-length serovar E DNA sequence of CT622.

SEQ ID NO:29 sets forth the DNA sequence for clone E3-B4-18 (CT1 positive) that contains a 1224 bp insert containing 4 ORFs. The complete ORF for CT772, and the partial ORFs of CT771, CT191, and CT190.

SEQ ID NO:30 sets forth the DNA sequence for the clone E9-E10-51 (CT10 positive) that contains an 883 bp insert containing two partial ORF, CT680 and CT679.

SEQ ID NO:31 sets forth the DNA sequence of the clone E9-D5-8 (CT10, CTCT1, CT4, and CT11 positive) that contains a 393 bp insert containing the partial ORF for CT680.

SEQ ID NO:32 sets forth the DNA sequence of the clone E7-B1-16 (CT10, CT3, CT5, CT11, CT13, and CHH037 positive) that contains a 2577 bp insert containing three ORFs, two full length ORFs for CT694 and CT695 and the third containing the N-terminal portion of CT969.

SEQ ID NO:33 sets forth the DNA sequence of the clone E9-G2-93 (CT10 positive) that contains a 554 bp insert containing a partial ORF for CT178.

SEQ ID NO:34 sets forth the DNA sequence of the clone E5-A8-85 (CT1 positive) that contains a 1433 bp insert containing two partial ORFs for CT875 and CT001.

SEQ ID NO:35 sets forth the DNA sequence of the clone E10-C6-45 (CT3 positive) that contains a 196 bp insert containing a partial ORF for CT827.

SEQ ID NO:36 sets forth the DNA sequence of the clone E7-H11-10 (CT3 positive) that contains a 1990 bp insert containing the partial ORFs of CT610 and CT613 and the complete ORFs of CT611 and CT612.

SEQ ID NO:37 sets forth the DNA sequence of the clone E2-F7-11 (CT3 and CT10 positive) that contains a 2093 bp insert. It contains a large region of CT609, a complete ORF for CT610 and a partial ORF for CT611.

SEQ ID NO:38 sets forth the DNA sequence of the clone E3-A3-31 (CT1 positive) that contains an 1834 bp insert containing a large region of CT622.

SEQ ID NO:39 sets forth the DNA sequence of the clone E1-G9-23 (CT3 positive) that contains an 1180 bp insert containing almost the entire ORF for CT798.

SEQ ID NO:40 sets forth the DNA sequence of the clone E4-D6-21 (CT3 positive) that contains a 1297 bp insert containing the partial ORFs of CT329 and CT327 and the complete ORF of CT328.

SEQ ID NO:41 sets forth the DNA sequence of the clone E3-F3-18 (CT1 positive) that contains an 1141 bp insert containing the partial ORF of CT871.

SEQ ID NO:42 sets forth the DNA sequence of the clone E10-B2-57 (CT10 positive) that contains an 822 bp insert containing the complete ORF of CT066.

5 SEQ ID NO:43 sets forth the DNA sequence of the clone E3-F3-7 (CT1 positive) that contains a 1643 bp insert containing the partial ORFs of CT869 and CT870.

SEQ ID NO:44 sets forth the DNA sequence of the clone E10-H8-1 (CT3 and CT10 positive) that contains an 1862 bp insert containing the partial ORFs of CT871 and CT872.

10 SEQ ID NO:45 sets forth the DNA sequence of the clone E3-D10-46 (CT1, CT3, CT4, CT11, and CT12 positive) that contains a 1666 bp insert containing the partial ORFs for CT770 and CT773 and the complete ORFs for CT771 and CT722.

SEQ ID NO:46 sets forth the DNA sequence of the clone E2-D8-19 (CT1 positive) that contains a 2010 bp insert containing partial ORFs, ORF3 and ORF6, and  
15 complete ORFs, ORF4 and ORF5.

SEQ ID NO:47 sets forth the DNA sequence of the clone E4-C3-40 (CT10 positive) that contains a 2044 bp insert containing the partial ORF for CT827 and a complete ORF for CT828.

SEQ ID NO:48 sets forth the DNA sequence of the clone E3-H6-10 (CT12 positive) that contains a 3743 bp insert containing the partial ORFs for CT223 and CT229 and the complete ORFs for CT224 and CT224, CT225, CT226, CT227, and CT228.

SEQ ID NO:49 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0454 of the *Chlamydia trachomatis* gene CT872.

SEQ ID NO:50 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0187, of the *Chlamydia trachomatis* gene CT133.  
25

SEQ ID NO:51 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0075 of the *Chlamydia trachomatis* gene CT321.

SEQ ID NO:52 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0074, of the *Chlamydia trachomatis* gene CT322.

30 SEQ ID NO:53 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0948, of the *Chlamydia trachomatis* gene CT798.

SEQ ID NO:54 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0985, of the *Chlamydia trachomatis* gene CT828.

SEQ ID NO:55 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0984, of the *Chlamydia trachomatis* gene CT827.

5 SEQ ID NO:56 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0062, of the *Chlamydia trachomatis* gene CT289.

SEQ ID NO:57 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn00065, of the *Chlamydia trachomatis* gene CT288.

10 SEQ ID NO:58 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0438, of the *Chlamydia trachomatis* gene CT287.

SEQ ID NO:59 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0963, of the *Chlamydia trachomatis* gene CT812.

SEQ ID NO:60 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0778, of the *Chlamydia trachomatis* gene CT603.

15 SEQ ID NO:61 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0503, of the *Chlamydia trachomatis* gene CT396.

SEQ ID NO:62 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn1016, of the *Chlamydia trachomatis* gene CT858.

20 SEQ ID NO:63 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0728, of the *Chlamydia trachomatis* gene CT622.

SEQ ID NO:64 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0557, of the *Chlamydia trachomatis* gene CT460.

SEQ ID NO:65 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0454, of the *Chlamydia trachomatis* gene CT872.

25 SEQ ID NO:66 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0187, of the *Chlamydia trachomatis* gene CT133.

SEQ ID NO:67 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0075, of the *Chlamydia trachomatis* gene CT321.

30 SEQ ID NO:68 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0074, of the *Chlamydia trachomatis* gene CT322.

SEQ ID NO:69 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0948, of the *Chlamydia trachomatis* gene CT798.

SEQ ID NO:70 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0985, of the *Chlamydia trachomatis* gene CT828.

5 SEQ ID NO:71 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0984, of the *Chlamydia trachomatis* gene CT827.

SEQ ID NO:72 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0062, of the *Chlamydia trachomatis* gene CT289.

10 SEQ ID NO:73 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0065, of the *Chlamydia trachomatis* gene CT288.

SEQ ID NO:74 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0438, of the *Chlamydia trachomatis* gene CT287.

SEQ ID NO:75 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0963, of the *Chlamydia trachomatis* gene CT812.

15 SEQ ID NO:76 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0778, of the *Chlamydia trachomatis* gene CT603.

SEQ ID NO:77 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn1016, of the *Chlamydia trachomatis* gene CT858.

20 SEQ ID NO:78 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0728, of the *Chlamydia trachomatis* gene CT622.

SEQ ID NO:79 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0557, of the *Chlamydia trachomatis* gene CT460.

SEQ ID NO:80 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT872.

25 SEQ ID NO:81 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT828.

SEQ ID NO:82 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT827.

30 SEQ ID NO:83 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT812.

SEQ ID NO:84 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT798.

SEQ ID NO:85 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT681 (MompF).

5           SEQ ID NO:86 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT603.

SEQ ID NO:87 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT460.

10          SEQ ID NO:88 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT322.

SEQ ID NO:89 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT321.

SEQ ID NO:90 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT289.

15          SEQ ID NO:91 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT288.

SEQ ID NO:92 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT287.

20          SEQ ID NO:93 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT133.

SEQ ID NO:94 sets forth the full-length serovar D DNA sequence of the *Chlamydia trachomatis* gene CT113.

SEQ ID NO:95 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT872.

25          SEQ ID NO:96 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT828.

SEQ ID NO:97 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT827.

30          SEQ ID NO:98 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT812.



SEQ ID NO:99 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT798.

SEQ ID NO:100 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT681.

5 SEQ ID NO:101 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT603.

SEQ ID NO:102 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT460.

10 SEQ ID NO:103 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT322.

SEQ ID NO:104 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT321.

SEQ ID NO:105 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT289.

15 SEQ ID NO:106 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT288.

SEQ ID NO:107 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT287.

20 SEQ ID NO:108 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT133.

SEQ ID NO:109 sets forth the full-length serovar D amino acid sequence of the *Chlamydia trachomatis* gene CT113.

SEQ ID NO:110 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0695, of the *Chlamydia trachomatis* gene CT681.

25 SEQ ID NO:111 sets forth the DNA sequence for the *Chlamydia pneumoniae* homologue, CPn0144, of the *Chlamydia trachomatis* gene CT113.

SEQ ID NO:112 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0695, of the *Chlamydia trachomatis* gene CT681.

30 SEQ ID NO:113 sets forth the amino acid sequence for the *Chlamydia pneumoniae* homologue, CPn0144, of the *Chlamydia trachomatis* gene CT113.

SEQ ID NO:114 sets forth the DNA sequence of the clone E7-B12-65 (CHH037 positive) that contains a 1179 bp insert containing complete ORF for 376.

SEQ ID NO:115 sets forth the DNA sequence of the clone E4-H9-83 (CHH037 positive) that contains the partial ORF for the heat shock protein GroEL  
5 (CT110).

SEQ ID NO:116 sets forth the DNA sequence of the clone E9-B10-52 (CHH037 positive) that contains the partial ORF for the the gene yscC (CT674).

SEQ ID NO:117 sets forth the DNA sequence of the clone E7-A7-79 (CHH037 positive) that contains the complete ORF for the histone like development gene  
10 hctA (CT743) and a partial ORF for the rRNA methyltransferase gene ygcA (CT742).

SEQ ID NO:118 sets forth the DNA sequence of the clone E2-D11-18 (CHH037 positive) that contains the partial ORF for hctA (CT743).

SEQ ID NO:119 sets forth the DNA sequence for the Chlamydia trachomatis serovar E hypothetical protein CT694.

15 SEQ ID NO:120 sets forth the DNA sequence for the Chlamydia trachomatis serovar E hypothetical protein CT695.

SEQ ID NO:121 sets forth the DNA sequence for the Chlamydia trachomatis serovar E L1 ribosomal protein.

20 SEQ ID NO:122 sets forth the amino acid sequence for the Chlamydia trachomatis serovar E hypothetical protein CT694.

SEQ ID NO:123 sets forth the amino acid sequence for the Chlamydia trachomatis serovar E hypothetical protein CT695.

SEQ ID NO:124 sets forth the amino acid sequence for the Chlamydia trachomatis serovar E L1 ribosomal protein.

25 SEQ ID NO:125 sets forth the DNA sequence of the clone E9-H6-15 (CT3 positive) that contains the partial ORF for the pmpB gene (CT413).

SEQ ID NO:126 sets forth the DNA sequence of the clone E3-D10-87 (CT1 positive) that contains the partial ORFs for the hypothetical genes CT388 and CT389.

30 SEQ ID NO:127 sets forth the DNA sequence of the clone E9-D6-43 (CT3 positive) that contains the partial ORF for the CT858.

SEQ ID NO:128 sets forth the DNA sequence of the clone E3-D10-4 (CT1 positive) that contains the partial ORF for pGP3-D, an ORF encoded on the plasmid pCHL1.

5 SEQ ID NO:129 sets forth the DNA sequence of the clone E3-G8-7 (CT1 positive) that contains the partial ORFs for the CT557 (LpdA) and CT558 (LipA).

SEQ ID NO:130 sets forth the DNA sequence of the clone E3-F11-32 (CT1 positive) that contains the partial ORF for pmpD (CT812).

10 SEQ ID NO:131 sets forth the DNA sequence of the clone E2-F8-5 (CT12 positive) that contains the complete ORF for the 15 kDa ORF (CT442) and a partial ORF for the 60kDa ORF (CT443).

SEQ ID NO:132 sets forth the DNA sequence of the clone E2-G4-39 (CT12 positive) that contains the partial ORF for the 60kDa ORF (CT443).

SEQ ID NO:133 sets forth the DNA sequence of the clone E9-D1-16 (CT10 positive) that contains the partial ORF for pmpH (CT872).

15 SEQ ID NO:134 sets forth the DNA sequence of the clone E3-F3-6 (CT1 positive) that contains the partial ORFs for the genes accB (CT123), L1 ribosomal (CT125) and S9 ribosomal (CT126).

SEQ ID NO:135 sets forth the DNA sequence of the clone E2-D4-70 (CT12 positive) that contains the partial ORF for the pmpC gene (CT414).

20 SEQ ID NO:136 sets forth the DNA sequence of the clone E5-A1-79 (CT1 positive) that contains the partial ORF for ydhO (CT127), a complete ORF for S9 ribosomal gene (CT126), a complete ORF for the L1 ribosomal gene (CT125) and a partial ORF for accC (CT124).

25 SEQ ID NO:137 sets forth the DNA sequence of the clone E1-F7-16 (CT12, CT3, and CT11 positive) that contains the partial ORF for the ftsH gene (CT841) and the entire ORF for the pnp gene (CT842).

SEQ ID NO:138 sets forth the DNA sequence of the clone E1-D8-62 (CT12 positive) that contains the partial ORFs for the ftsH gene (CT841) and for the pnp gene (CT842).

30 SEQ ID NO:139 sets forth the amino acid sequence for the serovar E protein CT875.

SEQ ID NO:140 sets forth the amino acid sequence for the serovar E protein CT622.

SEQ ID NO:141 sets forth the DNA sequence for the clone E8-C12-38, identified using the line CHH042 that contains the partial ORFs for sfhB (CT658) and  
5 CT659.

SEQ ID NO:142 sets forth the DNA sequence for the clone E1-D12-36, identified using the line CHH042 that contains the partial ORFs for mreB (CT709) (CT658) and pckA (CT710).

SEQ ID NO:143 sets forth the DNA sequence for the clone E8-D1-46,  
10 identified using the line CHH037 that contains the almost complete ORF for the pepA gene (CT045).

SEQ ID NO:144 sets forth the DNA sequence for the clone E10-A11-10, identified using the line CHH007 that contains the partial ORFs for yscU (CT091) and truB gene (CT094) as well as complete ORFs for ychF (CT092) and ribF (CT093).

15 SEQ ID NO:145 sets forth the DNA sequence for the clone E8-B12-80, identified using the line CHH037 that contains a partial ORF for the dag\_2 gene (CT735), a short fragment of the SET domain protein (CT737), as well as a complete ORF for ybcL (CT736).

SEQ ID NO:146 sets forth the DNA sequence for the clone E2-A8-70,  
20 identified using the line CHH037 that contains partial ORFs for the mutS gene (CT792) and the dag\_2 gene (CT735) as well as a complete ORF for the ybcL gene (CT736).

SEQ ID NO:147 sets forth the DNA sequence for the clone E10-C1-47, identified using the line CHH037 that contains the partial ORFs for the yaeI gene (CT461) and the prfB gene (CT459) as well as a complete ORF for the SWIB gene (CT460).

25 SEQ ID NO:148 sets forth the DNA sequence for the clone E8-G7-86, identified using the line CHH037 that contains partial ORFs for the mesJ gene (CT840) and the ftsH gene (CT841).

SEQ ID NO:149 sets forth the DNA sequence for the clone E3-E6-84, identified using the line CHH037 that contains partial ORFs for the pmpC gene (CT414)  
30 and the hypothetical protein CT611.

SEQ ID NO:150 sets for the DNA sequence for the clone E2-A11-49, identified using the patient line CHH042, that contains partial ORFs for the HAD superfamily (CT103) and the hypothetical protein CT105, as well as a complete ORF for *fabI* (CT104).

5           SEQ ID NO:151 sets for the DNA sequence for the clone E9-E6-4, identified using the patient line CHH042, it contains a complete ORF for the hypothetical protein CT659 and a partial ORF for *gyrA-2* (CT660).

          SEQ ID NO:152 sets for the DNA sequence for the clone E4-G8-49, identified using the patient line CHH042, it contains partial ORFs for the genes *pckA* (CT710) and *mreB* (CT709), as well as a partial ORF for the pGP2-D sequence derived from the plasmid.

          SEQ ID NO:153 sets for the DNA sequence for the clone E10-A8-16, identified using the patient line CHH042, it contains partial ORFs for the genes *rS3* (CT522) and *rL3* (CT528), as well as complete ORFs for the genes *rL22* (CT523), *rS19* (CT524), *rL2* (CT525), *rL23* (CT526) and *rL4* (CT527).

          SEQ ID NO:154 sets for the DNA sequence for the clone E10-F12-58, identified using the patient line CHH042, that contains partial ORFs for the genes *mhpA* (CT148), *rL16* (CT521), and *rL22* (CT523) as well as complete ORFs for the genes *rS3* (CT522), *rL22* (CT523) and *rS19* (CT524).

20           SEQ ID NO:155 sets for the DNA sequence for the clone E10-F12-42, identified using the patient line CHH042, that contains partial ORFs for the genes *rS3* (CT522) and *rL23* (CT526), as well as complete ORFs for the genes *rL22* (CT523), *rS19* (CT524) and *rL2* (CT525).

          SEQ ID NO:156 sets for the DNA sequence for the clone E2-C3-27, identified using the patient line CHH042, that contains partial ORFs for the genes *rL16* (CT521) and *rS19* (CT524), as well as complete ORFs for the genes *rS3* (CT522) and *rL22* (CT523).

          SEQ ID NO:157 sets forth the DNA sequence for the clone E2-A11-49, identified using the patient CHH037, that contains partial ORFs for the *ftsH* gene (CT841), pGP7-D and pGP5-D, as well as a complete ORF for pGP6-D.

## DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

As noted above, the present invention is generally directed to compositions and methods for the diagnosis and treatment of Chlamydial infection. In one aspect, the compositions of the subject invention include polypeptides that comprise at least one immunogenic portion of a *Chlamydia* antigen, or a variant thereof.

In specific embodiments, the subject invention discloses polypeptides comprising an immunogenic portion of a *Chlamydia* antigen, wherein the *Chlamydia* antigen comprises an amino acid sequence encoded by a polynucleotide molecule including a sequence selected from the group consisting of (a) nucleotide sequences recited in SEQ ID NO:1-48, 114-121, 125-138, and 141-157 (b) the complements of said nucleotide sequences, and (c) variants of such sequences.

## POLYNUCLEOTIDE COMPOSITIONS

As used herein, the terms "DNA segment" and "polynucleotide" refer to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding a polypeptide refers to a DNA segment that contains one or more coding sequences yet is substantially isolated away from, or purified free from, total genomic DNA of the species from which the DNA segment is obtained. Included within the terms "DNA segment" and "polynucleotide" are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phagemids, phage, viruses, and the like.

As will be understood by those skilled in the art, the DNA segments of this invention can include genomic sequences, extra-genomic and plasmid-encoded sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, peptides and the like. Such segments may be naturally isolated, or modified synthetically by the hand of man.

"Isolated," as used herein, means that a polynucleotide is substantially away from other coding sequences, and that the DNA segment does not contain large portions of unrelated coding DNA, such as large chromosomal fragments or other functional genes or polypeptide coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

As will be recognized by the skilled artisan, polynucleotides may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules include HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Polynucleotides may comprise a native Chlamydia sequence or may comprise a variant, or a biological or antigenic functional equivalent of such a sequence. Polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions, as further described below, preferably such that the immunogenicity of the encoded polypeptide is not diminished, relative to a native Chlamydia protein. The effect on the immunogenicity of the encoded polypeptide may generally be assessed as described herein. The term "variants" also encompasses homologous genes of xenogenic origin.

When comparing polynucleotide or polypeptide sequences, two sequences are said to be "identical" if the sequence of nucleotides or amino acids in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183,

Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of*  
 5 *Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA* 80:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.*  
 10 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining  
 15 percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides and polypeptides of the invention. Software for  
 20 performing BLAST analyses is publicly available through the National Center for Biotechnology Information. In one illustrative example, cumulative scores can be calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix can be used to calculate the cumulative score.  
 25 Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLASTN program (for  
 30 nucleotide sequences) uses as defaults a wordlength (W) of 11, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad.*



*Sci. USA* 89:10915) alignments, (B) of 50, expectation (E) of 10, M=5, N=-4 and a comparison of both strands.

Preferably, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein  
 5 the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid bases  
 10 or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

Therefore, the present invention encompasses polynucleotide and  
 15 polypeptide sequences having substantial identity to the sequences disclosed herein, for example those comprising at least 50% sequence identity, preferably at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% or higher, sequence identity compared to a polynucleotide or polypeptide sequence of this invention using the methods described herein, (e.g., BLAST analysis using standard parameters, as described  
 20 below). One skilled in this art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like.

In additional embodiments, the present invention provides isolated  
 25 polynucleotides and polypeptides comprising various lengths of contiguous stretches of sequence identical to or complementary to one or more of the sequences disclosed herein. For example, polynucleotides are provided by this invention that comprise at least about 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there  
 30 between. It will be readily understood that "intermediate lengths", in this context, means any length between the quoted values, such as 16, 17, 18, 19, *etc.*; 21, 22, 23, *etc.*; 30, 31,

32, *etc.*; 50, 51, 52, 53, *etc.*; 100, 101, 102, 103, *etc.*; 150, 151, 152, 153, *etc.*; including all integers through 200-500; 500-1,000, and the like.

The polynucleotides of the present invention, or fragments thereof, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, illustrative DNA segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

In other embodiments, the present invention is directed to polynucleotides that are capable of hybridizing under moderately stringent conditions to a polynucleotide sequence provided herein, or a fragment thereof, or a complementary sequence thereof. Hybridization techniques are well known in the art of molecular biology. For purposes of illustration, suitable moderately stringent conditions for testing the hybridization of a polynucleotide of this invention with other polynucleotides include prewashing in a solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5 X SSC, overnight; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS.

Moreover, it will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not, have an

altered structure or function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

#### PROBES AND PRIMERS

5           In other embodiments of the present invention, the polynucleotide sequences provided herein can be advantageously used as probes or primers for nucleic acid hybridization. As such, it is contemplated that nucleic acid segments that comprise a sequence region of at least about 15 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 15 nucleotide long contiguous sequence disclosed  
10   herein will find particular utility. Longer contiguous identical or complementary sequences, *e.g.*, those of about 20, 30, 40, 50, 100, 200, 500, 1000 (including all intermediate lengths) and even up to full length sequences will also be of use in certain embodiments.

          The ability of such nucleic acid probes to specifically hybridize to a sequence  
15   of interest will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are also envisioned, such as the use of the sequence information for the preparation of mutant species primers, or primers for use in preparing other genetic constructions.

          Polynucleotide molecules having sequence regions consisting of contiguous  
20   nucleotide stretches of 10-14, 15-20, 30, 50, or even of 100-200 nucleotides or so (including intermediate lengths as well), identical or complementary to a polynucleotide sequence disclosed herein, are particularly contemplated as hybridization probes for use in, *e.g.*, Southern and Northern blotting. This would allow a gene product, or fragment thereof, to be analyzed, both in diverse cell types and also in various bacterial cells. The  
25   total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the contiguous complementary region may be varied, such as between about 15 and about 100 nucleotides, but larger contiguous complementarity stretches may be used, according to the  
30   length complementary sequences one wishes to detect.

The use of a hybridization probe of about 15-25 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 15 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 25 contiguous nucleotides, or even longer where desired.

Hybridization probes may be selected from any portion of any of the sequences disclosed herein. All that is required is to review the sequence set forth in SEQ ID NO:1-48, 114-121, 125-138, and 141-157, or to any continuous portion of the sequence, from about 15-25 nucleotides in length up to and including the full length sequence, that one wishes to utilize as a probe or primer. The choice of probe and primer sequences may be governed by various factors. For example, one may wish to employ primers from towards the termini of the total sequence.

Small polynucleotide segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR<sup>TM</sup> technology of U. S. Patent 4,683,202 (incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

The nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of the entire gene or gene fragments of interest. Depending on the application envisioned, one will typically desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, one will select relatively low salt and/or high temperature conditions, such as provided by a salt concentration of from about 0.02 M to about 0.15 M salt at temperatures of from about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the

probe and the template or target strand, and would be particularly suitable for isolating related sequences.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template, less stringent (reduced stringency) hybridization conditions will typically be needed in order to allow formation of the heteroduplex. In these circumstances, one may desire to employ salt conditions such as those of from about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

15

#### POLYNUCLEOTIDE IDENTIFICATION AND CHARACTERIZATION

Polynucleotides may be identified, prepared and/or manipulated using any of a variety of well established techniques. For example, a polynucleotide may be identified, by screening a microarray of cDNAs for Chlamydia expression. Such screens may be performed, for example, using a Synteni microarray (Palo Alto, CA) according to the manufacturer's instructions (and essentially as described by Schena et al., *Proc. Natl. Acad. Sci. USA* 93:10614-10619, 1996 and Heller et al., *Proc. Natl. Acad. Sci. USA* 94:2150-2155, 1997). Alternatively, polynucleotides may be amplified from cDNA prepared from cells expressing the proteins described herein. Such polynucleotides may be amplified via polymerase chain reaction (PCR). For this approach, sequence-specific primers may be designed based on the sequences provided herein, and may be purchased or synthesized.

An amplified portion of a polynucleotide of the present invention may be used to isolate a full length gene from a suitable library (e.g., Chlamydia cDNA library) using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification.

Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

For hybridization techniques, a partial sequence may be labeled (*e.g.*, by  
5 nick-translation or end-labeling with  $^{32}\text{P}$ ) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (*see* Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). Hybridizing colonies or plaques are selected  
10 and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a series of  
15 deletion clones. The resulting overlapping sequences can then be assembled into a single contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

Alternatively, there are numerous amplification techniques for obtaining a full length coding sequence from a partial cDNA sequence. Within such techniques,  
20 amplification is generally performed via PCR. Any of a variety of commercially available kits may be used to perform the amplification step. Primers may be designed using, for example, software well known in the art. Primers are preferably 22-30 nucleotides in length, have a GC content of at least 50% and anneal to the target sequence at temperatures of about 68°C to 72°C. The amplified region may be sequenced as described above, and  
25 overlapping sequences assembled into a contiguous sequence.

One such amplification technique is inverse PCR (*see* Triglia et al., *Nucl. Acids Res.* 16:8186, 1988), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within  
30 an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region.

The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as  
5 "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom et al., *PCR Methods Applic. 1*:111-19, 1991) and walking PCR (Parker et al., *Nucl. Acids. Res. 19*:3055-60, 1991). Other methods  
10 employing amplification may also be employed to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (e.g., NCBI BLAST searches), and such ESTs may be used to  
15 generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

#### POLYNUCLEOTIDE EXPRESSION IN HOST CELLS

In other embodiments of the invention, polynucleotide sequences or  
20 fragments thereof which encode polypeptides of the invention, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to  
25 clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant  
30 RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

Moreover, the polynucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For example, DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. In addition, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

10 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of polypeptide activity, it may be useful to encode a chimeric protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the polypeptide-encoding sequence and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

Sequences encoding a desired polypeptide may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 215-223, Horn, T. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of a polypeptide, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al. (1995) *Science* 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

25 A newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) *Proteins, Structures and Molecular Principles*, WH Freeman and Co., New York, N.Y.) or other comparable techniques available in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure). Additionally, the amino acid sequence of a polypeptide, or any part thereof, may be altered



during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a desired polypeptide, the nucleotide sequences encoding the polypeptide, or functional equivalents, may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York. N.Y.

A variety of expression vector/host systems may be utilized to contain and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The "control elements" or "regulatory sequences" present in an expression vector are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the PBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSORT1 plasmid (Gibco BRL, Gaithersburg, MD) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is

necessary to generate a cell line that contains multiple copies of the sequence encoding a polypeptide, vectors based on SV40 or EBV may be advantageously used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large quantities are needed, for example for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of  $\beta$ -galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509); and the like. pGEX Vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) *Methods Enzymol.* 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311. Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl.*

*Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196).

An insect system may also be used to express a polypeptide of interest. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which the polypeptide of interest may be expressed (Engelhard, E. K. et al. (1994) *Proc. Natl. Acad. Sci.* 91 :3224-3227).

In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan, J. and Shenk, T. (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted,

exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression  
5 may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired  
10 fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms  
15 for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a polynucleotide of interest may be transformed using expression vectors which may contain viral origins of  
20 replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced  
25 sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) *Cell* 11:223-32) and adenine phosphoribosyltransferase (Lowy, I.  
30 et al. (1990) *Cell* 22:817-23) genes which can be employed in tk.sup.- or aprt.sup.- cells,

respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci.* 77:3567-70); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al (1981) *J. Mol. Biol.* 150:1-14); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) *Proc. Natl. Acad. Sci.* 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al. (1995) *Methods Mol. Biol.* 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence, recombinant cells containing sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a polypeptide-encoding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain and express a desired polynucleotide sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies specific for the product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell

sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on a given polypeptide may be preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological  
5 Methods, a Laboratory Manual, APS Press, St Paul. Minn.) and Maddox, D. E. et al. (1983; *J. Exp. Med.* 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to  
10 polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled  
15 nucleotides. These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with a polynucleotide sequence of interest may be  
20 cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the invention may be designed to contain signal sequences which direct secretion of the encoded polypeptide  
25 through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals,  
30 protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle,

Wash.). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, Calif.) between the purification domain and the encoded polypeptide may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a polypeptide of interest and a  
5 nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography) as described in Porath, J. et al. (1992, *Prot. Exp. Purif.* 3:263-281) while the enterokinase cleavage site provides a means for purifying the desired polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins  
10 is provided in Kroll, D. J. et al. (1993; *DNA Cell Biol.* 12:441-453).

In addition to recombinant production methods, polypeptides of the invention, and fragments thereof, may be produced by direct peptide synthesis using solid-phase techniques (Merrifield J. (1963) *J. Am. Chem. Soc.* 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be  
15 achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Alternatively, various fragments may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

#### SITE-SPECIFIC MUTAGENESIS

20 Site-specific mutagenesis is a technique useful in the preparation of individual peptides, or biologically functional equivalent polypeptides, through specific mutagenesis of the underlying polynucleotides that encode them. The technique, well-known to those of skill in the art, further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations,  
25 by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being  
30 traversed. Mutations may be employed in a selected polynucleotide sequence to improve, alter, decrease, modify, or otherwise change the properties of the polynucleotide itself,

and/or alter the properties, activity, composition, stability, or primary sequence of the encoded polypeptide.

In certain embodiments of the present invention, the inventors contemplate the mutagenesis of the disclosed polynucleotide sequences to alter one or more properties of the encoded polypeptide, such as the antigenicity of a polypeptide vaccine. The techniques of site-specific mutagenesis are well-known in the art, and are widely used to create variants of both polypeptides and polynucleotides. For example, site-specific mutagenesis is often used to alter a specific portion of a DNA molecule. In such embodiments, a primer comprising typically about 14 to about 25 nucleotides or so in length is employed, with about 5 to about 10 residues on both sides of the junction of the sequence being altered.

As will be appreciated by those of skill in the art, site-specific mutagenesis techniques have often employed a phage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially-available and their use is generally well-known to those skilled in the art. Double-stranded plasmids are also routinely employed in site directed mutagenesis that eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of a double-stranded vector that includes within its sequence a DNA sequence that encodes the desired peptide. An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected peptide-encoding DNA segments using site-directed mutagenesis provides a means of producing potentially useful



species and is not meant to be limiting as there are other ways in which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants. Specific details regarding these methods and protocols are found in the teachings of Maloy *et al.*, 1994; Segal, 1976; Prokop and Bajpai, 1991; Kubby, 1994; and Maniatis *et al.*, 1982, each incorporated herein by reference, for that purpose.

As used herein, the term "oligonucleotide directed mutagenesis procedure" refers to template-dependent processes and vector-mediated propagation which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable signal, such as amplification. As used herein, the term "oligonucleotide directed mutagenesis procedure" is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template dependent process refers to nucleic acid synthesis of an RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known rules of complementary base pairing (see, for example, Watson, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by U. S. Patent No. 4,237,224, specifically incorporated herein by reference in its entirety.

#### POLYNUCLEOTIDE AMPLIFICATION TECHNIQUES

A number of template dependent processes are available to amplify the target sequences of interest present in a sample. One of the best known amplification methods is the polymerase chain reaction (PCR<sup>TM</sup>) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, each of which is incorporated herein by reference in its entirety. Briefly, in PCR<sup>TM</sup>, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the target sequence. An excess of deoxynucleoside triphosphates is added to a reaction mixture along with a DNA polymerase (e.g., *Taq* polymerase). If the target sequence is present in a sample, the

primers will bind to the target and the polymerase will cause the primers to be extended along the target sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the target to form reaction products, excess primers will bind to the target and to the reaction product and the process is repeated. Preferably reverse transcription and PCR<sup>TM</sup> amplification procedure may be performed in order to quantify the amount of mRNA amplified. Polymerase chain reaction methodologies are well known in the art.

Another method for amplification is the ligase chain reaction (referred to as LCR), disclosed in Eur. Pat. Appl. Publ. No. 320,308 (specifically incorporated herein by reference in its entirety). In LCR, two complementary probe pairs are prepared, and in the presence of the target sequence, each pair will bind to opposite complementary strands of the target such that they abut. In the presence of a ligase, the two probe pairs will link to form a single unit. By temperature cycling, as in PCR<sup>TM</sup>, bound ligated units dissociate from the target and then serve as "target sequences" for ligation of excess probe pairs. U.S. Patent No. 4,883,750, incorporated herein by reference in its entirety, describes an alternative method of amplification similar to LCR for binding probe pairs to a target sequence.

Qbeta Replicase, described in PCT Intl. Pat. Appl. Publ. No. PCT/US87/00880, incorporated herein by reference in its entirety, may also be used as still another amplification method in the present invention. In this method, a replicative sequence of RNA that has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence that can then be detected.

An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[ $\alpha$ -thio]triphosphates in one strand of a restriction site (Walker *et al.*, 1992, incorporated herein by reference in its entirety), may also be useful in the amplification of nucleic acids in the present invention.

Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, *i.e.* nick translation. A similar method, called Repair Chain

Reaction (RCR) is another method of amplification which may be useful in the present invention and is involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases can be added as biotinylated derivatives for easy detection. A similar  
5 approach is used in SDA.

Sequences can also be detected using a cyclic probe reaction (CPR). In CPR, a probe having a 3' and 5' sequences of non-target DNA and an internal or "middle" sequence of the target protein specific RNA is hybridized to DNA which is present in a sample. Upon hybridization, the reaction is treated with RNaseH, and the products of the  
10 probe are identified as distinctive products by generating a signal that is released after digestion. The original template is annealed to another cycling probe and the reaction is repeated. Thus, CPR involves amplifying a signal generated by hybridization of a probe to a target gene specific expressed nucleic acid.

Still other amplification methods described in Great Britain Pat. Appl. No. 2  
15 202 328, and in PCT Intl. Pat. Appl. Publ. No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCR-like, template and enzyme dependent synthesis. The primers may be modified by labeling with a capture moiety (e.g., biotin) and/or a detector moiety (e.g., enzyme). In the latter application, an  
20 excess of labeled probes is added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe signals the presence of the target sequence.

Other nucleic acid amplification procedures include transcription-based  
25 amplification systems (TAS) (Kwoh *et al.*, 1989; PCT Intl. Pat. Appl. Publ. No. WO 88/10315, incorporated herein by reference in its entirety), including nucleic acid sequence based amplification (NASBA) and 3SR. In NASBA, the nucleic acids can be prepared for amplification by standard phenol/chloroform extraction, heat denaturation of a sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or  
30 guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer that has sequences specific to the target sequence. Following

polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat-denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target-specific primer, followed by polymerization. The double stranded DNA molecules are then multiply transcribed by a polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNAs are reverse transcribed into DNA, and transcribed once again with a polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target-specific sequences.

Eur. Pat. Appl. Publ. No. 329,822, incorporated herein by reference in its entirety, disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention. The ssRNA is a first template for a first primer oligonucleotide, which is elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from resulting DNA:RNA duplex by the action of ribonuclease H (RNase H, an RNase specific for RNA in a duplex with either DNA or RNA). The resultant ssDNA is a second template for a second primer, which also includes the sequences of an RNA polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to its template. This primer is then extended by DNA polymerase (exemplified by the large "Klenow" fragment of *E. coli* DNA polymerase I), resulting as a double-stranded DNA ("dsDNA") molecule, having a sequence identical to that of the original RNA between the primers and having additionally, at one end, a promoter sequence. This promoter sequence can be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies can then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification can be done isothermally without addition of enzymes at each cycle. Because of the cyclical nature of this process, the starting sequence can be chosen to be in the form of either DNA or RNA.

PCT Intl. Pat. Appl. Publ. No. WO 89/06700, incorporated herein by reference in its entirety, disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic; i.e. new templates are not produced from the resultant RNA transcripts. Other

amplification methods include "RACE" (Frohman, 1990), and "one-sided PCR" (Ohara, 1989) which are well-known to those of skill in the art.

Methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby  
5 amplifying the di-oligonucleotide (Wu and Dean, 1996, incorporated herein by reference in its entirety), may also be used in the amplification of DNA sequences of the present invention.

#### BIOLOGICAL FUNCTIONAL EQUIVALENTS

10 Modification and changes may be made in the structure of the polynucleotides and polypeptides of the present invention and still obtain a functional molecule that encodes a polypeptide with desirable characteristics. As mentioned above, it is often desirable to introduce one or more mutations into a specific polynucleotide sequence. In certain circumstances, the resulting encoded polypeptide sequence is altered  
15 by this mutation, or in other cases, the sequence of the polypeptide is unchanged by one or more mutations in the encoding polynucleotide.

When it is desirable to alter the amino acid sequence of a polypeptide to create an equivalent, or even an improved, second-generation molecule, the amino acid changes may be achieved by changing one or more of the codons of the encoding DNA  
20 sequence, according to Table 1.

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines  
25 that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides without appreciable loss of  
30 their biological utility or activity.

TABLE I

Amino Acids			Codons						
Alanine	Ala	A	GCA	GCC	GCG	GCU			
Cysteine	Cys	C	UGC	UGU					
Aspartic acid	Asp	D	GAC	GAU					
Glutamic acid	Glu	E	GAA	GAG					
Phenylalanine	Phe	F	UUC	UUU					
Glycine	Gly	G	GGA	GGC	GGG	GGU			
Histidine	His	H	CAC	CAU					
Isoleucine	Ile	I	AUA	AUC	AUU				
Lysine	Lys	K	AAA	AAG					
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU	
Methionine	Met	M	AUG						
Asparagine	Asn	N	AAC	AAU					
Proline	Pro	P	CCA	CCC	CCG	CCU			
Glutamine	Gln	Q	CAA	CAG					
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU	
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU	
Threonine	Thr	T	ACA	ACC	ACG	ACU			
Valine	Val	V	GUA	GUC	GUG	GUU			
Tryptophan	Trp	W	UGG						
Tyrosine	Tyr	Y	UAC	UAU					

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporated herein by reference). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like. Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, 1982). These values are: isoleucine (+4.5); valine (+4.2); leucine

(+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (−0.4); threonine (−0.7); serine (−0.8); tryptophan (−0.9); tyrosine (−1.3); proline (−1.6); histidine (−3.2); glutamate (−3.5); glutamine (−3.5); aspartate (−3.5); asparagine (−3.5); lysine (−3.9); and arginine (−4.5).

5 It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.* still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within  $\pm 2$  is preferred, those within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even  
10 more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U. S. Patent 4,554,101 (specifically incorporated herein by reference in its entirety), states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

15 As detailed in U. S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0  $\pm$  1); glutamate (+3.0  $\pm$  1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (−0.4); proline (−0.5  $\pm$  1); alanine (−0.5); histidine (−0.5); cysteine (−1.0); methionine (−1.3); valine (−1.5); leucine (−1.8); isoleucine (−1.8); tyrosine (−2.3);  
20 phenylalanine (−2.5); tryptophan (−3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within  $\pm 2$  is preferred, those within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly  
25 preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take various of the foregoing characteristics into consideration are well known to those of  
30 skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

In addition, any polynucleotide may be further modified to increase stability *in vivo*. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends; the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages in the backbone; and/or the inclusion of nontraditional bases such as inosine, queosine and wybutosine, as well as acetyl- methyl-, thio- and other modified forms of adenine, cytidine, guanine, thymine and uridine.

#### IN VIVO POLYNUCLEOTIDE DELIVERY TECHNIQUES

In additional embodiments, genetic constructs comprising one or more of the polynucleotides of the invention are introduced into cells *in vivo*. This may be achieved using any of a variety of well known approaches, several of which are outlined below for the purpose of illustration.

##### 1. ADENOVIRUS

One of the preferred methods for *in vivo* delivery of one or more nucleic acid sequences involves the use of an adenovirus expression vector. "Adenovirus expression vector" is meant to include those constructs containing adenovirus sequences sufficient to (a) support packaging of the construct and (b) to express a polynucleotide that has been cloned therein in a sense or antisense orientation. Of course, in the context of an antisense construct, expression does not require that the gene product be synthesized.

The expression vector comprises a genetically engineered form of an adenovirus. Knowledge of the genetic organization of adenovirus, a 36 kb, linear, double-stranded DNA virus, allows substitution of large pieces of adenoviral DNA with foreign sequences up to 7 kb (Grunhaus and Horwitz, 1992). In contrast to retrovirus, the adenoviral infection of host cells does not result in chromosomal integration because adenoviral DNA can replicate in an episomal manner without potential genotoxicity. Also, adenoviruses are structurally stable, and no genome rearrangement has been detected after extensive amplification. Adenovirus can infect virtually all epithelial cells regardless of their cell cycle stage. So far, adenoviral infection appears to be linked only to mild disease such as acute respiratory disease in humans.



Adenovirus is particularly suitable for use as a gene transfer vector because of its mid-sized genome, ease of manipulation, high titer, wide target-cell range and high infectivity. Both ends of the viral genome contain 100-200 base pair inverted repeats (ITRs), which are *cis* elements necessary for viral DNA replication and packaging. The early (E) and late (L) regions of the genome contain different transcription units that are divided by the onset of viral DNA replication. The E1 region (E1A and E1B) encodes proteins responsible for the regulation of transcription of the viral genome and a few cellular genes. The expression of the E2 region (E2A and E2B) results in the synthesis of the proteins for viral DNA replication. These proteins are involved in DNA replication, late gene expression and host cell shut-off (Renan, 1990). The products of the late genes, including the majority of the viral capsid proteins, are expressed only after significant processing of a single primary transcript issued by the major late promoter (MLP). The MLP, (located at 16.8 m.u.) is particularly efficient during the late phase of infection, and all the mRNA's issued from this promoter possess a 5'-tripartite leader (TPL) sequence which makes them preferred mRNA's for translation.

In a current system, recombinant adenovirus is generated from homologous recombination between shuttle vector and provirus vector. Due to the possible recombination between two proviral vectors, wild-type adenovirus may be generated from this process. Therefore, it is critical to isolate a single clone of virus from an individual plaque and examine its genomic structure.

Generation and propagation of the current adenovirus vectors, which are replication deficient, depend on a unique helper cell line, designated 293, which was transformed from human embryonic kidney cells by Ad5 DNA fragments and constitutively expresses E1 proteins (Graham *et al.*, 1977). Since the E3 region is dispensable from the adenovirus genome (Jones and Shenk, 1978), the current adenovirus vectors, with the help of 293 cells, carry foreign DNA in either the E1, the D3 or both regions (Graham and Prevec, 1991). In nature, adenovirus can package approximately 105% of the wild-type genome (Ghosh-Choudhury *et al.*, 1987), providing capacity for about 2 extra kB of DNA. Combined with the approximately 5.5 kB of DNA that is replaceable in the E1 and E3 regions, the maximum capacity of the current adenovirus vector is under 7.5 kB, or about 15% of the total length of the vector. More than 80% of

the adenovirus viral genome remains in the vector backbone and is the source of vector-borne cytotoxicity. Also, the replication deficiency of the E1-deleted virus is incomplete. For example, leakage of viral gene expression has been observed with the currently available vectors at high multiplicities of infection (MOI) (Mulligan, 1993).

5           Helper cell lines may be derived from human cells such as human embryonic kidney cells, muscle cells, hematopoietic cells or other human embryonic mesenchymal or epithelial cells. Alternatively, the helper cells may be derived from the cells of other mammalian species that are permissive for human adenovirus. Such cells include, *e.g.*, Vero cells or other monkey embryonic mesenchymal or epithelial cells. As  
10       stated above, the currently preferred helper cell line is 293.

          Recently, Racher *et al.* (1995) disclosed improved methods for culturing 293 cells and propagating adenovirus. In one format, natural cell aggregates are grown by inoculating individual cells into 1 liter siliconized spinner flasks (Techne, Cambridge, UK) containing 100-200 ml of medium. Following stirring at 40 rpm, the cell viability is  
15       estimated with trypan blue. In another format, Fibra-Cel microcarriers (Bibby Sterlin, Stone, UK) (5 g/l) is employed as follows. A cell inoculum, resuspended in 5 ml of medium, is added to the carrier (50 ml) in a 250 ml Erlenmeyer flask and left stationary, with occasional agitation, for 1 to 4 h. The medium is then replaced with 50 ml of fresh medium and shaking initiated. For virus production, cells are allowed to grow to about  
20       80% confluence, after which time the medium is replaced (to 25% of the final volume) and adenovirus added at an MOI of 0.05. Cultures are left stationary overnight, following which the volume is increased to 100% and shaking commenced for another 72 h.

          Other than the requirement that the adenovirus vector be replication defective, or at least conditionally defective, the nature of the adenovirus vector is not  
25       believed to be crucial to the successful practice of the invention. The adenovirus may be of any of the 42 different known serotypes or subgroups A-F. Adenovirus type 5 of subgroup C is the preferred starting material in order to obtain a conditional replication-defective adenovirus vector for use in the present invention, since Adenovirus type 5 is a human adenovirus about which a great deal of biochemical and genetic information is known, and  
30       it has historically been used for most constructions employing adenovirus as a vector.

As stated above, the typical vector according to the present invention is replication defective and will not have an adenovirus E1 region. Thus, it will be most convenient to introduce the polynucleotide encoding the gene of interest at the position from which the E1-coding sequences have been removed. However, the position of  
 5 insertion of the construct within the adenovirus sequences is not critical to the invention. The polynucleotide encoding the gene of interest may also be inserted in lieu of the deleted E3 region in E3 replacement vectors as described by Karlsson *et al.* (1986) or in the E4 region where a helper cell line or helper virus complements the E4 defect.

Adenovirus is easy to grow and manipulate and exhibits broad host range *in vitro* and *in vivo*. This group of viruses can be obtained in high titers, *e.g.*,  $10^9$ - $10^{11}$  plaque-forming units per ml, and they are highly infective. The life cycle of adenovirus does not require integration into the host cell genome. The foreign genes delivered by adenovirus vectors are episomal and, therefore, have low genotoxicity to host cells. No side effects have been reported in studies of vaccination with wild-type adenovirus (Couch *et al.*, 1963;  
 10 Top *et al.*, 1971), demonstrating their safety and therapeutic potential as *in vivo* gene transfer vectors.

Adenovirus vectors have been used in eukaryotic gene expression (Levrero *et al.*, 1991; Gomez-Foix *et al.*, 1992) and vaccine development (Grunhaus and Horwitz, 1992; Graham and Prevec, 1992). Recently, animal studies suggested that recombinant  
 20 adenovirus could be used for gene therapy (Stratford-Perricaudet and Perricaudet, 1991; Stratford-Perricaudet *et al.*, 1990; Rich *et al.*, 1993). Studies in administering recombinant adenovirus to different tissues include trachea instillation (Rosenfeld *et al.*, 1991; Rosenfeld *et al.*, 1992), muscle injection (Ragot *et al.*, 1993), peripheral intravenous injections (Herz and Gerard, 1993) and stereotactic inoculation into the brain (Le Gal La  
 25 Salle *et al.*, 1993).

## 2. RETROVIRUSES

The retroviruses are a group of single-stranded RNA viruses characterized by an ability to convert their RNA to double-stranded DNA in infected cells by a process of  
 30 reverse-transcription (Coffin, 1990). The resulting DNA then stably integrates into cellular chromosomes as a provirus and directs synthesis of viral proteins. The integration results

in the retention of the viral gene sequences in the recipient cell and its descendants. The retroviral genome contains three genes, gag, pol, and env that code for capsid proteins, polymerase enzyme, and envelope components, respectively. A sequence found upstream from the gag gene contains a signal for packaging of the genome into virions. Two long  
5 terminal repeat (LTR) sequences are present at the 5' and 3' ends of the viral genome. These contain strong promoter and enhancer sequences and are also required for integration in the host cell genome (Coffin, 1990).

In order to construct a retroviral vector, a nucleic acid encoding one or more oligonucleotide or polynucleotide sequences of interest is inserted into the viral genome in  
10 the place of certain viral sequences to produce a virus that is replication-defective. In order to produce virions, a packaging cell line containing the gag, pol, and env genes but without the LTR and packaging components is constructed (Mann *et al.*, 1983). When a recombinant plasmid containing a cDNA, together with the retroviral LTR and packaging sequences is introduced into this cell line (by calcium phosphate precipitation for example),  
15 the packaging sequence allows the RNA transcript of the recombinant plasmid to be packaged into viral particles, which are then secreted into the culture media (Nicolas and Rubenstein, 1988; Temin, 1986; Mann *et al.*, 1983). The media containing the recombinant retroviruses is then collected, optionally concentrated, and used for gene transfer. Retroviral vectors are able to infect a broad variety of cell types. However,  
20 integration and stable expression require the division of host cells (Paskind *et al.*, 1975).

A novel approach designed to allow specific targeting of retrovirus vectors was recently developed based on the chemical modification of a retrovirus by the chemical addition of lactose residues to the viral envelope. This modification could permit the specific infection of hepatocytes *via* sialoglycoprotein receptors.

25 A different approach to targeting of recombinant retroviruses was designed in which biotinylated antibodies against a retroviral envelope protein and against a specific cell receptor were used. The antibodies were coupled *via* the biotin components by using streptavidin (Roux *et al.*, 1989). Using antibodies against major histocompatibility complex class I and class II antigens, they demonstrated the infection of a variety of human  
30 cells that bore those surface antigens with an ecotropic virus *in vitro* (Roux *et al.*, 1989).

### 3. ADENO-ASSOCIATED VIRUSES

AAV (Ridgeway, 1988; Hermonat and Muzycska, 1984) is a parovirus, discovered as a contamination of adenoviral stocks. It is a ubiquitous virus (antibodies are present in 85% of the US human population) that has not been linked to any disease. It is also classified as a dependovirus, because its replications is dependent on the presence of a helper virus, such as adenovirus. Five serotypes have been isolated, of which AAV-2 is the best characterized. AAV has a single-stranded linear DNA that is encapsidated into capsid proteins VP1, VP2 and VP3 to form an icosahedral virion of 20 to 24 nm in diameter (Muzyczka and McLaughlin, 1988).

The AAV DNA is approximately 4.7 kilobases long. It contains two open reading frames and is flanked by two ITRs (FIG. 2). There are two major genes in the AAV genome: *rep* and *cap*. The *rep* gene codes for proteins responsible for viral replications, whereas *cap* codes for capsid protein VP1-3. Each ITR forms a T-shaped hairpin structure. These terminal repeats are the only essential *cis* components of the AAV for chromosomal integration. Therefore, the AAV can be used as a vector with all viral coding sequences removed and replaced by the cassette of genes for delivery. Three viral promoters have been identified and named p5, p19, and p40, according to their map position. Transcription from p5 and p19 results in production of *rep* proteins, and transcription from p40 produces the capsid proteins (Hermonat and Muzyczka, 1984).

There are several factors that prompted researchers to study the possibility of using rAAV as an expression vector. One is that the requirements for delivering a gene to integrate into the host chromosome are surprisingly few. It is necessary to have the 145-bp ITRs, which are only 6% of the AAV genome. This leaves room in the vector to assemble a 4.5-kb DNA insertion. While this carrying capacity may prevent the AAV from delivering large genes, it is amply suited for delivering the antisense constructs of the present invention.

AAV is also a good choice of delivery vehicles due to its safety. There is a relatively complicated rescue mechanism: not only wild type adenovirus but also AAV genes are required to mobilize rAAV. Likewise, AAV is not pathogenic and not associated with any disease. The removal of viral coding sequences minimizes immune reactions to viral gene expression, and therefore, rAAV does not evoke an inflammatory response.

#### 4. OTHER VIRAL VECTORS AS EXPRESSION CONSTRUCTS

Other viral vectors may be employed as expression constructs in the present invention for the delivery of oligonucleotide or polynucleotide sequences to a host cell. Vectors derived from viruses such as vaccinia virus (Ridgeway, 1988; Coupar *et al.*, 1988),  
5 lentiviruses, polio viruses and herpes viruses may be employed. They offer several attractive features for various mammalian cells (Friedmann, 1989; Ridgeway, 1988; Coupar *et al.*, 1988; Horwich *et al.*, 1990).

With the recent recognition of defective hepatitis B viruses, new insight was gained into the structure-function relationship of different viral sequences. *In vitro* studies  
10 showed that the virus could retain the ability for helper-dependent packaging and reverse transcription despite the deletion of up to 80% of its genome (Horwich *et al.*, 1990). This suggested that large portions of the genome could be replaced with foreign genetic material. The hepatotropism and persistence (integration) were particularly attractive properties for liver-directed gene transfer. Chang *et al.* (1991) introduced the  
15 chloramphenicol acetyltransferase (CAT) gene into duck hepatitis B virus genome in the place of the polymerase, surface, and pre-surface coding sequences. It was cotransfected with wild-type virus into an avian hepatoma cell line. Culture media containing high titers of the recombinant virus were used to infect primary duckling hepatocytes. Stable CAT gene expression was detected for at least 24 days after transfection (Chang *et al.*, 1991).

20

#### 5. NON-VIRAL VECTORS

In order to effect expression of the oligonucleotide or polynucleotide sequences of the present invention, the expression construct must be delivered into a cell. This delivery may be accomplished *in vitro*, as in laboratory procedures for transforming  
25 cells lines, or *in vivo* or *ex vivo*, as in the treatment of certain disease states. As described above, one preferred mechanism for delivery is *via* viral infection where the expression construct is encapsulated in an infectious viral particle.

Once the expression construct has been delivered into the cell the nucleic acid encoding the desired oligonucleotide or polynucleotide sequences may be positioned  
30 and expressed at different sites. In certain embodiments, the nucleic acid encoding the construct may be stably integrated into the genome of the cell. This integration may be in

the specific location and orientation *via* homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle. How the expression construct is delivered to a cell and where in the cell the nucleic acid remains is dependent on the type of expression construct employed.

In certain embodiments of the invention, the expression construct comprising one or more oligonucleotide or polynucleotide sequences may simply consist of naked recombinant DNA or plasmids. Transfer of the construct may be performed by any of the methods mentioned above which physically or chemically permeabilize the cell membrane. This is particularly applicable for transfer *in vitro* but it may be applied to *in vivo* use as well. Dubensky *et al.* (1984) successfully injected polyomavirus DNA in the form of calcium phosphate precipitates into liver and spleen of adult and newborn mice demonstrating active viral replication and acute infection. Benvenisty and Reshef (1986) also demonstrated that direct intraperitoneal injection of calcium phosphate-precipitated plasmids results in expression of the transfected genes. It is envisioned that DNA encoding a gene of interest may also be transferred in a similar manner *in vivo* and express the gene product.

Another embodiment of the invention for transferring a naked DNA expression construct into cells may involve particle bombardment. This method depends on the ability to accelerate DNA-coated microprojectiles to a high velocity allowing them to pierce cell membranes and enter cells without killing them (Klein *et al.*, 1987). Several devices for accelerating small particles have been developed. One such device relies on a high voltage discharge to generate an electrical current, which in turn provides the motive force (Yang *et al.*, 1990). The microprojectiles used have consisted of biologically inert substances such as tungsten or gold beads.

Selected organs including the liver, skin, and muscle tissue of rats and mice have been bombarded *in vivo* (Yang *et al.*, 1990; Zelenin *et al.*, 1991). This may require surgical exposure of the tissue or cells, to eliminate any intervening tissue between the gun

and the target organ, *i.e.* *ex vivo* treatment. Again, DNA encoding a particular gene may be delivered *via* this method and still be incorporated by the present invention.

#### ANTISENSE OLIGONUCLEOTIDES

5           The end result of the flow of genetic information is the synthesis of protein. DNA is transcribed by polymerases into messenger RNA and translated on the ribosome to yield a folded, functional protein. Thus there are several steps along the route where protein synthesis can be inhibited. The native DNA segment coding for a polypeptide described herein, as all such mammalian DNA strands, has two strands: a sense strand and  
10   an antisense strand held together by hydrogen bonding. The messenger RNA coding for polypeptide has the same nucleotide sequence as the sense DNA strand except that the DNA thymidine is replaced by uridine. Thus, synthetic antisense nucleotide sequences will bind to a mRNA and inhibit expression of the protein encoded by that mRNA.

          The targeting of antisense oligonucleotides to mRNA is thus one mechanism  
15   to shut down protein synthesis, and, consequently, represents a powerful and targeted therapeutic approach. For example, the synthesis of polygalacturonase and the muscarine type 2 acetylcholine receptor are inhibited by antisense oligonucleotides directed to their respective mRNA sequences (U. S. Patent 5,739,119 and U. S. Patent 5,759,829, each specifically incorporated herein by reference in its entirety). Further, examples of antisense  
20   inhibition have been demonstrated with the nuclear protein cyclin, the multiple drug resistance gene (MDG1), ICAM-1, E-selectin, STK-1, striatal GABA<sub>A</sub> receptor and human EGF (Jaskulski *et al.*, 1988; Vasanthakumar and Ahmed, 1989; Peris *et al.*, 1998; U. S. Patent 5,801,154; U. S. Patent 5,789,573; U. S. Patent 5,718,709 and U. S. Patent 5,610,288, each specifically incorporated herein by reference in its entirety). Antisense  
25   constructs have also been described that inhibit and can be used to treat a variety of abnormal cellular proliferations, *e.g.* cancer (U. S. Patent 5,747,470; U. S. Patent 5,591,317 and U. S. Patent 5,783,683, each specifically incorporated herein by reference in its entirety).

          Therefore, in exemplary embodiments, the invention provides  
30   oligonucleotide sequences that comprise all, or a portion of, any sequence that is capable of specifically binding to polynucleotide sequence described herein, or a complement thereof.



In one embodiment, the antisense oligonucleotides comprise DNA or derivatives thereof. In another embodiment, the oligonucleotides comprise RNA or derivatives thereof. In a third embodiment, the oligonucleotides are modified DNAs comprising a phosphorothioated modified backbone. In a fourth embodiment, the oligonucleotide sequences comprise peptide nucleic acids or derivatives thereof. In each case, preferred compositions comprise a sequence region that is complementary, and more preferably substantially-complementary, and even more preferably, completely complementary to one or more portions of polynucleotides disclosed herein.

Selection of antisense compositions specific for a given gene sequence is based upon analysis of the chosen target sequence (*i.e.* in these illustrative examples the rat and human sequences) and determination of secondary structure,  $T_m$ , binding energy, relative stability, and antisense compositions were selected based upon their relative inability to form dimers, hairpins, or other secondary structures that would reduce or prohibit specific binding to the target mRNA in a host cell.

Highly preferred target regions of the mRNA, are those which are at or near the AUG translation initiation codon, and those sequences which were substantially complementary to 5' regions of the mRNA. These secondary structure analyses and target site selection considerations were performed using v.4 of the OLIGO primer analysis software (Rychlik, 1997) and the BLASTN 2.0.5 algorithm software (Altschul *et al.*, 1997).

The use of an antisense delivery method employing a short peptide vector, termed MPG (27 residues), is also contemplated. The MPG peptide contains a hydrophobic domain derived from the fusion sequence of HIV gp41 and a hydrophilic domain from the nuclear localization sequence of SV40 T-antigen (Morris *et al.*, 1997). It has been demonstrated that several molecules of the MPG peptide coat the antisense oligonucleotides and can be delivered into cultured mammalian cells in less than 1 hour with relatively high efficiency (90%). Further, the interaction with MPG strongly increases both the stability of the oligonucleotide to nuclease and the ability to cross the plasma membrane (Morris *et al.*, 1997).

## RIBOZYMES

Although proteins traditionally have been used for catalysis of nucleic acids, another class of macromolecules has emerged as useful in this endeavor. Ribozymes are RNA-protein complexes that cleave nucleic acids in a site-specific fashion. Ribozymes have specific catalytic domains that possess endonuclease activity (Kim and Cech, 1987; Gerlach *et al.*, 1987; Forster and Symons, 1987). For example, a large number of ribozymes accelerate phosphoester transfer reactions with a high degree of specificity, often cleaving only one of several phosphoesters in an oligonucleotide substrate (Cech *et al.*, 1981; Michel and Westhof, 1990; Reinhold-Hurek and Shub, 1992). This specificity has been attributed to the requirement that the substrate bind via specific base-pairing interactions to the internal guide sequence ("IGS") of the ribozyme prior to chemical reaction.

Ribozyme catalysis has primarily been observed as part of sequence-specific cleavage/ligation reactions involving nucleic acids (Joyce, 1989; Cech *et al.*, 1981). For example, U. S. Patent No. 5,354,855 (specifically incorporated herein by reference) reports that certain ribozymes can act as endonucleases with a sequence specificity greater than that of known ribonucleases and approaching that of the DNA restriction enzymes. Thus, sequence-specific ribozyme-mediated inhibition of gene expression may be particularly suited to therapeutic applications (Scanlon *et al.*, 1991; Sarver *et al.*, 1990). Recently, it was reported that ribozymes elicited genetic changes in some cells lines to which they were applied; the altered genes included the oncogenes *H-ras*, *c-fos* and genes of HIV. Most of this work involved the modification of a target mRNA, based on a specific mutant codon that is cleaved by a specific ribozyme.

Six basic varieties of naturally-occurring enzymatic RNAs are known presently. Each can catalyze the hydrolysis of RNA phosphodiester bonds *in trans* (and thus can cleave other RNA molecules) under physiological conditions. In general, enzymatic nucleic acids act by first binding to a target RNA. Such binding occurs through the target binding portion of a enzymatic nucleic acid which is held in close proximity to an enzymatic portion of the molecule that acts to cleave the target RNA. Thus, the enzymatic nucleic acid first recognizes and then binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cut the target RNA.

Strategic cleavage of such a target RNA will destroy its ability to direct synthesis of an encoded protein. After an enzymatic nucleic acid has bound and cleaved its RNA target, it is released from that RNA to search for another target and can repeatedly bind and cleave new targets.

5           The enzymatic nature of a ribozyme is advantageous over many technologies, such as antisense technology (where a nucleic acid molecule simply binds to a nucleic acid target to block its translation) since the concentration of ribozyme necessary to affect a therapeutic treatment is lower than that of an antisense oligonucleotide. This advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme molecule is able to cleave many molecules of target RNA. In addition, the ribozyme is a highly specific inhibitor, with the specificity of inhibition depending not only on the base pairing mechanism of binding to the target RNA, but also on the mechanism of target RNA cleavage. Single mismatches, or base-substitutions, near the site of cleavage can completely eliminate catalytic activity of a ribozyme. Similar mismatches in antisense molecules do not prevent their action (Woolf *et al.*, 1992). Thus, the specificity of action of a ribozyme is greater than that of an antisense oligonucleotide binding the same RNA site.

          The enzymatic nucleic acid molecule may be formed in a hammerhead, hairpin, a hepatitis  $\delta$  virus, group I intron or RNaseP RNA (in association with an RNA guide sequence) or Neurospora VS RNA motif. Examples of hammerhead motifs are described by Rossi *et al.* (1992). Examples of hairpin motifs are described by Hampel *et al.* (Eur. Pat. Appl. Publ. No. EP 0360257), Hampel and Tritz (1989), Hampel *et al.* (1990) and U. S. Patent 5,631,359 (specifically incorporated herein by reference). An example of the hepatitis  $\delta$  virus motif is described by Perrotta and Been (1992); an example of the RNaseP motif is described by Guerrier-Takada *et al.* (1983); Neurospora VS RNA ribozyme motif is described by Collins (Saville and Collins, 1990; Saville and Collins, 1991; Collins and Olive, 1993); and an example of the Group I intron is described in (U. S. Patent 4,987,071, specifically incorporated herein by reference). All that is important in an enzymatic nucleic acid molecule of this invention is that it has a specific substrate binding site which is complementary to one or more of the target gene RNA regions, and that it have nucleotide sequences within or surrounding that substrate binding site which impart

an RNA cleaving activity to the molecule. Thus the ribozyme constructs need not be limited to specific motifs mentioned herein.

In certain embodiments, it may be important to produce enzymatic cleaving agents which exhibit a high degree of specificity for the RNA of a desired target, such as one of the sequences disclosed herein. The enzymatic nucleic acid molecule is preferably targeted to a highly conserved sequence region of a target mRNA. Such enzymatic nucleic acid molecules can be delivered exogenously to specific cells as required. Alternatively, the ribozymes can be expressed from DNA or RNA vectors that are delivered to specific cells.

Small enzymatic nucleic acid motifs (e.g., of the hammerhead or the hairpin structure) may also be used for exogenous delivery. The simple structure of these molecules increases the ability of the enzymatic nucleic acid to invade targeted regions of the mRNA structure. Alternatively, catalytic RNA molecules can be expressed within cells from eukaryotic promoters (e.g., Scanlon *et al.*, 1991; Kashani-Sabet *et al.*, 1992; Dropulic *et al.*, 1992; Weerasinghe *et al.*, 1991; Ojwang *et al.*, 1992; Chen *et al.*, 1992; Sarver *et al.*, 1990). Those skilled in the art realize that any ribozyme can be expressed in eukaryotic cells from the appropriate DNA vector. The activity of such ribozymes can be augmented by their release from the primary transcript by a second ribozyme (Int. Pat. Appl. Publ. No. WO 93/23569, and Int. Pat. Appl. Publ. No. WO 94/02595, both hereby incorporated by reference; Ohkawa *et al.*, 1992; Taira *et al.*, 1991; and Ventura *et al.*, 1993).

Ribozymes may be added directly, or can be complexed with cationic lipids, lipid complexes, packaged within liposomes, or otherwise delivered to target cells. The RNA or RNA complexes can be locally administered to relevant tissues *ex vivo*, or *in vivo* through injection, aerosol inhalation, infusion pump or stent, with or without their incorporation in biopolymers.

Ribozymes may be designed as described in Int. Pat. Appl. Publ. No. WO 93/23569 and Int. Pat. Appl. Publ. No. WO 94/02595, each specifically incorporated herein by reference) and synthesized to be tested *in vitro* and *in vivo*, as described. Such ribozymes can also be optimized for delivery. While specific examples are provided, those in the art will recognize that equivalent RNA targets in other species can be utilized when necessary.

Hammerhead or hairpin ribozymes may be individually analyzed by computer folding (Jaeger *et al.*, 1989) to assess whether the ribozyme sequences fold into the appropriate secondary structure. Those ribozymes with unfavorable intramolecular interactions between the binding arms and the catalytic core are eliminated from  
5 consideration. Varying binding arm lengths can be chosen to optimize activity. Generally, at least 5 or so bases on each arm are able to bind to, or otherwise interact with, the target RNA.

Ribozymes of the hammerhead or hairpin motif may be designed to anneal to various sites in the mRNA message, and can be chemically synthesized. The method of  
10 synthesis used follows the procedure for normal RNA synthesis as described in Usman *et al.* (1987) and in Scaringe *et al.* (1990) and makes use of common nucleic acid protecting and coupling groups, such as dimethoxytrityl at the 5'-end, and phosphoramidites at the 3'-end. Average stepwise coupling yields are typically >98%. Hairpin ribozymes may be synthesized in two parts and annealed to reconstruct an active  
15 ribozyme (Chowrira and Burke, 1992). Ribozymes may be modified extensively to enhance stability by modification with nuclease resistant groups, for example, 2'-amino, 2'-C-allyl, 2'-fluoro, 2'-o-methyl, 2'-H (for a review see *e.g.*, Usman and Cedergren, 1992). Ribozymes may be purified by gel electrophoresis using general methods or by high pressure liquid chromatography and resuspended in water.

20 Ribozyme activity can be optimized by altering the length of the ribozyme binding arms, or chemically synthesizing ribozymes with modifications that prevent their degradation by serum ribonucleases (see *e.g.*, Int. Pat. Appl. Publ. No. WO 92/07065; Perrault *et al.*, 1990; Pieken *et al.*, 1991; Usman and Cedergren, 1992; Int. Pat. Appl. Publ. No. WO 93/15187; Int. Pat. Appl. Publ. No. WO 91/03162; Eur. Pat. Appl. Publ.  
25 No. 92110298.4; U. S. Patent 5,334,711; and Int. Pat. Appl. Publ. No. WO 94/13688, which describe various chemical modifications that can be made to the sugar moieties of enzymatic RNA molecules), modifications which enhance their efficacy in cells, and removal of stem II bases to shorten RNA synthesis times and reduce chemical requirements.

30 Sullivan *et al.* (Int. Pat. Appl. Publ. No. WO 94/02595) describes the general methods for delivery of enzymatic RNA molecules. Ribozymes may be

administered to cells by a variety of methods known to those familiar to the art, including, but not restricted to, encapsulation in liposomes, by iontophoresis, or by incorporation into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, and bioadhesive microspheres. For some indications, ribozymes may be directly delivered *ex vivo* to cells or tissues with or without the aforementioned vehicles. Alternatively, the RNA/vehicle combination may be locally delivered by direct inhalation, by direct injection or by use of a catheter, infusion pump or stent. Other routes of delivery include, but are not limited to, intravascular, intramuscular, subcutaneous or joint injection, aerosol inhalation, oral (tablet or pill form), topical, systemic, ocular, intraperitoneal and/or intrathecal delivery. More detailed descriptions of ribozyme delivery and administration are provided in Int. Pat. Appl. Publ. No. WO 94/02595 and Int. Pat. Appl. Publ. No. WO 93/23569, each specifically incorporated herein by reference.

Another means of accumulating high concentrations of a ribozyme(s) within cells is to incorporate the ribozyme-encoding sequences into a DNA expression vector. Transcription of the ribozyme sequences are driven from a promoter for eukaryotic RNA polymerase I (pol I), RNA polymerase II (pol II), or RNA polymerase III (pol III). Transcripts from pol II or pol III promoters will be expressed at high levels in all cells; the levels of a given pol II promoter in a given cell type will depend on the nature of the gene regulatory sequences (enhancers, silencers, *etc.*) present nearby. Prokaryotic RNA polymerase promoters may also be used, providing that the prokaryotic RNA polymerase enzyme is expressed in the appropriate cells (Elroy-Stein and Moss, 1990; Gao and Huang, 1993; Lieber *et al.*, 1993; Zhou *et al.*, 1990). Ribozymes expressed from such promoters can function in mammalian cells (*e.g.* Kashani-Saber *et al.*, 1992; Ojwang *et al.*, 1992; Chen *et al.*, 1992; Yu *et al.*, 1993; L'Huillier *et al.*, 1992; Lisiewicz *et al.*, 1993). Such transcription units can be incorporated into a variety of vectors for introduction into mammalian cells, including but not restricted to, plasmid DNA vectors, viral DNA vectors (such as adenovirus or adeno-associated vectors), or viral RNA vectors (such as retroviral, semliki forest virus, sindbis virus vectors).

Ribozymes may be used as diagnostic tools to examine genetic drift and mutations within diseased cells. They can also be used to assess levels of the target RNA molecule. The close relationship between ribozyme activity and the structure of the target

RNA allows the detection of mutations in any region of the molecule which alters the base-pairing and three-dimensional structure of the target RNA. By using multiple ribozymes, one may map nucleotide changes which are important to RNA structure and function *in vitro*, as well as in cells and tissues. Cleavage of target RNAs with ribozymes may be used to inhibit gene expression and define the role (essentially) of specified gene products in the progression of disease. In this manner, other genetic targets may be defined as important mediators of the disease. These studies will lead to better treatment of the disease progression by affording the possibility of combinational therapies (e.g., multiple ribozymes targeted to different genes, ribozymes coupled with known small molecule inhibitors, or intermittent treatment with combinations of ribozymes and/or other chemical or biological molecules). Other *in vitro* uses of ribozymes are well known in the art, and include detection of the presence of mRNA associated with an IL-5 related condition. Such RNA is detected by determining the presence of a cleavage product after treatment with a ribozyme using standard methodology.

#### PEPTIDE NUCLEIC ACIDS

In certain embodiments, the inventors contemplate the use of peptide nucleic acids (PNAs) in the practice of the methods of the invention. PNA is a DNA mimic in which the nucleobases are attached to a pseudopeptide backbone (Good and Nielsen, 1997). PNA is able to be utilized in a number methods that traditionally have used RNA or DNA. Often PNA sequences perform better in techniques than the corresponding RNA or DNA sequences and have utilities that are not inherent to RNA or DNA. A review of PNA including methods of making, characteristics of, and methods of using, is provided by Corey (1997) and is incorporated herein by reference. As such, in certain embodiments, one may prepare PNA sequences that are complementary to one or more portions of the ACE mRNA sequence, and such PNA compositions may be used to regulate, alter, decrease, or reduce the translation of ACE-specific mRNA, and thereby alter the level of ACE activity in a host cell to which such PNA compositions have been administered.

PNAs have 2-aminoethyl-glycine linkages replacing the normal phosphodiester backbone of DNA (Nielsen *et al.*, 1991; Hanvey *et al.*, 1992; Hyrup and Nielsen, 1996; Neilsen, 1996). This chemistry has three important consequences: firstly,

in contrast to DNA or phosphorothioate oligonucleotides, PNAs are neutral molecules; secondly, PNAs are achiral, which avoids the need to develop a stereoselective synthesis; and thirdly, PNA synthesis uses standard Boc (Dueholm *et al.*, 1994) or Fmoc (Thomson *et al.*, 1995) protocols for solid-phase peptide synthesis, although other methods, including a  
5 modified Merrifield method, have been used (Christensen *et al.*, 1995).

PNA monomers or ready-made oligomers are commercially available from PerSeptive Biosystems (Framingham, MA). PNA syntheses by either Boc or Fmoc protocols are straightforward using manual or automated protocols (Norton *et al.*, 1995). The manual protocol lends itself to the production of chemically modified PNAs or the  
10 simultaneous synthesis of families of closely related PNAs.

As with peptide synthesis, the success of a particular PNA synthesis will depend on the properties of the chosen sequence. For example, while in theory PNAs can incorporate any combination of nucleotide bases, the presence of adjacent purines can lead to deletions of one or more residues in the product. In expectation of this difficulty, it is  
15 suggested that, in producing PNAs with adjacent purines, one should repeat the coupling of residues likely to be added inefficiently. This should be followed by the purification of PNAs by reverse-phase high-pressure liquid chromatography (Norton *et al.*, 1995) providing yields and purity of product similar to those observed during the synthesis of peptides.

20 Modifications of PNAs for a given application may be accomplished by coupling amino acids during solid-phase synthesis or by attaching compounds that contain a carboxylic acid group to the exposed N-terminal amine. Alternatively, PNAs can be modified after synthesis by coupling to an introduced lysine or cysteine. The ease with which PNAs can be modified facilitates optimization for better solubility or for specific  
25 functional requirements. Once synthesized, the identity of PNAs and their derivatives can be confirmed by mass spectrometry. Several studies have made and utilized modifications of PNAs (Norton *et al.*, 1995; Haaima *et al.*, 1996; Stetsenko *et al.*, 1996; Petersen *et al.*, 1995; Ulmann *et al.*, 1996; Koch *et al.*, 1995; Orum *et al.*, 1995; Footer *et al.*, 1996; Griffith *et al.*, 1995; Kremsky *et al.*, 1996; Pardridge *et al.*, 1995; Boffa *et al.*, 1995;  
30 Landsdorp *et al.*, 1996; Gambacorti-Passerini *et al.*, 1996; Armitage *et al.*, 1997; Seeger *et al.*, 1997; Ruskowski *et al.*, 1997). U.S. Patent No. 5,700,922 discusses PNA-DNA-PNA



chimeric molecules and their uses in diagnostics, modulating protein in organisms, and treatment of conditions susceptible to therapeutics.

In contrast to DNA and RNA, which contain negatively charged linkages, the PNA backbone is neutral. In spite of this dramatic alteration, PNAs recognize  
 5 complementary DNA and RNA by Watson-Crick pairing (Egholm *et al.*, 1993), validating the initial modeling by Nielsen *et al.* (1991). PNAs lack 3' to 5' polarity and can bind in either parallel or antiparallel fashion, with the antiparallel mode being preferred (Egholm *et al.*, 1993).

Hybridization of DNA oligonucleotides to DNA and RNA is destabilized by  
 10 electrostatic repulsion between the negatively charged phosphate backbones of the complementary strands. By contrast, the absence of charge repulsion in PNA-DNA or PNA-RNA duplexes increases the melting temperature ( $T_m$ ) and reduces the dependence of  $T_m$  on the concentration of mono- or divalent cations (Nielsen *et al.*, 1991). The enhanced rate and affinity of hybridization are significant because they are responsible for the  
 15 surprising ability of PNAs to perform strand invasion of complementary sequences within relaxed double-stranded DNA. In addition, the efficient hybridization at inverted repeats suggests that PNAs can recognize secondary structure effectively within double-stranded DNA. Enhanced recognition also occurs with PNAs immobilized on surfaces, and Wang *et al.* have shown that support-bound PNAs can be used to detect hybridization events (Wang  
 20 *et al.*, 1996).

One might expect that tight binding of PNAs to complementary sequences would also increase binding to similar (but not identical) sequences, reducing the sequence specificity of PNA recognition. As with DNA hybridization, however, selective recognition can be achieved by balancing oligomer length and incubation temperature.  
 25 Moreover, selective hybridization of PNAs is encouraged by PNA-DNA hybridization being less tolerant of base mismatches than DNA-DNA hybridization. For example, a single mismatch within a 16 bp PNA-DNA duplex can reduce the  $T_m$  by up to 15°C (Egholm *et al.*, 1993). This high level of discrimination has allowed the development of several PNA-based strategies for the analysis of point mutations (Wang *et al.*, 1996;  
 30 Carlsson *et al.*, 1996; Thiede *et al.*, 1996; Webb and Hurskainen, 1996; Perry-O'Keefe *et al.*, 1996).

High-affinity binding provides clear advantages for molecular recognition and the development of new applications for PNAs. For example, 11-13 nucleotide PNAs inhibit the activity of telomerase, a ribonucleo-protein that extends telomere ends using an essential RNA template, while the analogous DNA oligomers do not (Norton *et al.*, 1996).

5 Neutral PNAs are more hydrophobic than analogous DNA oligomers, and this can lead to difficulty solubilizing them at neutral pH, especially if the PNAs have a high purine content or if they have the potential to form secondary structures. Their solubility can be enhanced by attaching one or more positive charges to the PNA termini (Nielsen *et al.*, 1991).

10 Findings by Allfrey and colleagues suggest that strand invasion will occur spontaneously at sequences within chromosomal DNA (Boffa *et al.*, 1995; Boffa *et al.*, 1996). These studies targeted PNAs to triplet repeats of the nucleotides CAG and used this recognition to purify transcriptionally active DNA (Boffa *et al.*, 1995) and to inhibit transcription (Boffa *et al.*, 1996). This result suggests that if PNAs can be delivered within  
15 cells then they will have the potential to be general sequence-specific regulators of gene expression. Studies and reviews concerning the use of PNAs as antisense and anti-gene agents include Nielsen *et al.* (1993b), Hanvey *et al.* (1992), and Good and Nielsen (1997). Koppelhus *et al.* (1997) have used PNAs to inhibit HIV-1 inverse transcription, showing that PNAs may be used for antiviral therapies.

20 Methods of characterizing the antisense binding properties of PNAs are discussed in Rose (1993) and Jensen *et al.* (1997). Rose uses capillary gel electrophoresis to determine binding of PNAs to their complementary oligonucleotide, measuring the relative binding kinetics and stoichiometry. Similar types of measurements were made by Jensen *et al.* using BIAcore™ technology.

25 Other applications of PNAs include use in DNA strand invasion (Nielsen *et al.*, 1991), antisense inhibition (Hanvey *et al.*, 1992), mutational analysis (Orum *et al.*, 1993), enhancers of transcription (Mollegaard *et al.*, 1994), nucleic acid purification (Orum *et al.*, 1995), isolation of transcriptionally active genes (Boffa *et al.*, 1995), blocking of transcription factor binding (Vickers *et al.*, 1995), genome cleavage (Veselkov *et al.*,  
30 1996), biosensors (Wang *et al.*, 1996), *in situ* hybridization (Thisted *et al.*, 1996), and in a alternative to Southern blotting (Perry-O'Keefe, 1996).

## POLYPEPTIDE COMPOSITIONS AND USES

The present invention, in other aspects, provides polypeptide compositions. Generally, a polypeptide of the invention will be an isolated polypeptide (or an epitope, variant, or active fragment thereof) derived from a mammalian species. Preferably, the polypeptide is encoded by a polynucleotide sequence disclosed herein or a sequence which hybridizes under moderately stringent conditions to a polynucleotide sequence disclosed herein. Alternatively, the polypeptide may be defined as a polypeptide which comprises a contiguous amino acid sequence from an amino acid sequence disclosed herein, or which polypeptide comprises an entire amino acid sequence disclosed herein.

Likewise, a polypeptide composition of the present invention is understood to comprise one or more polypeptides that are capable of eliciting antibodies that are immunologically reactive with one or more polypeptides encoded by one or more contiguous nucleic acid sequences contained in SEQ ID NO:1-48, 114-121, 125-138 and 141-157, or to active fragments, or to variants thereof, or to one or more nucleic acid sequences which hybridize to one or more of these sequences under conditions of moderate to high stringency.

As used herein, an active fragment of a polypeptide includes a whole or a portion of a polypeptide which is modified by conventional techniques, *e.g.*, mutagenesis, or by addition, deletion, or substitution, but which active fragment exhibits substantially the same structure function, antigenicity, etc., as a polypeptide as described herein.

In certain illustrative embodiments, the polypeptides of the invention will comprise at least an immunogenic portion of a Chlamydia protein or a variant thereof, as described herein. Proteins that are Chlamydia proteins generally also react detectably within an immunoassay (such as an ELISA) with antisera from a patient with a Chlamydial infection. Polypeptides as described herein may be of any length. Additional sequences derived from the native protein and/or heterologous sequences may be present, and such sequences may (but need not) possess further immunogenic or antigenic properties.

An "immunogenic portion," as used herein is a portion of a protein that is recognized (*i.e.*, specifically bound) by a B-cell and/or T-cell surface antigen receptor. Such immunogenic portions generally comprise at least 5 amino acid residues, more

preferably at least 10, and still more preferably at least 20 amino acid residues of a Chlamydia protein or a variant thereof. Certain preferred immunogenic portions include peptides in which an N-terminal leader sequence and/or transmembrane domain have been deleted. Other preferred immunogenic portions may contain a small N- and/or C-terminal  
 5 deletion (e.g., 1-30 amino acids, preferably 5-15 amino acids), relative to the mature protein.

Immunogenic portions may generally be identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (Raven Press, 1993) and references cited therein. Such techniques include screening  
 10 polypeptides for the ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen (i.e., they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well known techniques. An  
 15 immunogenic portion of a native Chlamydia protein is a portion that reacts with such antisera and/or T-cells at a level that is not substantially less than the reactivity of the full length polypeptide (e.g., in an ELISA and/or T-cell reactivity assay). Such immunogenic portions may react within such assays at a level that is similar to or greater than the reactivity of the full length polypeptide. Such screens may generally be performed using  
 20 methods well known to those of ordinary skill in the art, such as those described in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. For example, a polypeptide may be immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example, <sup>125</sup>I-  
 25 labeled Protein A.

As noted above, a composition may comprise a variant of a native Chlamydia protein. A polypeptide "variant," as used herein, is a polypeptide that differs from a native Chlamydia protein in one or more substitutions, deletions, additions and/or insertions, such that the immunogenicity of the polypeptide is not substantially diminished.  
 30 In other words, the ability of a variant to react with antigen-specific antisera may be enhanced or unchanged, relative to the native protein, or may be diminished by less than

50%, and preferably less than 20%, relative to the native protein. Such variants may generally be identified by modifying one of the above polypeptide sequences and evaluating the reactivity of the modified polypeptide with antigen-specific antibodies or antisera as described herein. Preferred variants include those in which one or more  
5 portions, such as an N-terminal leader sequence or transmembrane domain, have been removed. Other preferred variants include variants in which a small portion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids) has been removed from the N- and/or C-terminal of the mature protein.

Polypeptide variants encompassed by the present invention include those  
10 exhibiting at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more identity (determined as described above) to the polypeptides disclosed herein.

Preferably, a variant contains conservative substitutions. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has  
15 similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. Amino acid substitutions may generally be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues. For example, negatively charged amino acids include aspartic acid and  
20 glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine and valine; glycine and alanine; asparagine and glutamine; and serine, threonine, phenylalanine and tyrosine. Other groups of amino acids that may represent conservative changes include: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val,  
25 ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. A variant may also, or alternatively, contain nonconservative changes. In a preferred embodiment, variant polypeptides differ from a native sequence by substitution, deletion or addition of five amino acids or fewer. Variants may also (or alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenicity,  
30 secondary structure and hydropathic nature of the polypeptide.

As noted above, polypeptides may comprise a signal (or leader) sequence at the N-terminal end of the protein, which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (e.g., poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

Polypeptides may be prepared using any of a variety of well known techniques. Recombinant polypeptides encoded by DNA sequences as described above may be readily prepared from the DNA sequences using any of a variety of expression vectors known to those of ordinary skill in the art. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast, and higher eukaryotic cells, such as mammalian cells and plant cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO. Supernatants from suitable host/vector systems which secrete recombinant protein or polypeptide into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant polypeptide.

Portions and other variants having less than about 100 amino acids, and generally less than about 50 amino acids, may also be generated by synthetic means, using techniques well known to those of ordinary skill in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, *J. Am. Chem. Soc.* 85:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the manufacturer's instructions.

Within certain specific embodiments, a polypeptide may be a fusion protein that comprises multiple polypeptides as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence, such as a known Chlamydia protein. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the protein or to enable the protein to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the protein.

Fusion proteins may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion protein is expressed as a recombinant protein, allowing the production of increased levels, relative to a non-fused protein, in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion protein that retains the biological activity of both component polypeptides.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea

et al., *Gene* 40:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons required to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

Fusion proteins are also provided. Such proteins comprise a polypeptide as described herein together with an unrelated immunogenic protein. Preferably the immunogenic protein is capable of eliciting a recall response. Examples of such proteins include tetanus, tuberculosis and hepatitis proteins (*see*, for example, Stoute et al. *New Engl. J. Med.*, 336:86-91, 1997).

Within preferred embodiments, an immunological fusion partner is derived from protein D, a surface protein of the gram-negative bacterium *Haemophilus influenza B* (WO 91/18926). Preferably, a protein D derivative comprises approximately the first third of the protein (*e.g.*, the first N-terminal 100-110 amino acids), and a protein D derivative may be lipidated. Within certain preferred embodiments, the first 109 residues of a Lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells. Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemagglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the *LytA* gene; *Gene* 43:265-292, 1986). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-



terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino terminus has been described (see *Biotechnology* 10:795-798, 1992). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion protein. A repeat portion is found in the C-terminal region starting at residue 178. A particularly preferred repeat portion incorporates residues 188-305.

In general, polypeptides (including fusion proteins) and polynucleotides as described herein are isolated. An "isolated" polypeptide or polynucleotide is one that is removed from its original environment. For example, a naturally-occurring protein is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure. A polynucleotide is considered to be isolated if, for example, it is cloned into a vector that is not a part of the natural environment.

#### ILLUSTRATIVE THERAPEUTIC COMPOSITIONS AND USES

In another aspect, the present invention provides methods for using one or more of the above polypeptides or fusion proteins (or polynucleotides encoding such polypeptides or fusion proteins) to induce protective immunity against Chlamydial infection in a patient. As used herein, a "patient" refers to any warm-blooded animal, preferably a human. A patient may be afflicted with a disease, or may be free of detectable disease and/or infection. In other words, protective immunity may be induced to prevent or treat Chlamydial infection.

In this aspect, the polypeptide, fusion protein or polynucleotide molecule is generally present within a pharmaceutical composition or a vaccine. Pharmaceutical compositions may comprise one or more polypeptides, each of which may contain one or more of the above sequences (or variants thereof), and a physiologically acceptable carrier. Vaccines may comprise one or more of the above polypeptides and an immunostimulant, such as an adjuvant or a liposome (into which the polypeptide is incorporated). Such

pharmaceutical compositions and vaccines may also contain other *Chlamydia* antigens, either incorporated into a combination polypeptide or present within a separate polypeptide.

Alternatively, a vaccine may contain polynucleotides encoding one or more polypeptides or fusion proteins as described above, such that the polypeptide is generated *in situ*. In such vaccines, the polynucleotides may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacterial and viral expression systems. Appropriate nucleic acid expression systems contain the necessary polynucleotide sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface. In a preferred embodiment, the polynucleotides may be introduced using a viral expression system (*e.g.*, vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective) virus. Techniques for incorporating polynucleotides into such expression systems are well known to those of ordinary skill in the art. The polynucleotides may also be administered as "naked" plasmid vectors as described, for example, in Ulmer et al., *Science* 259:1745-1749, 1993 and reviewed by Cohen, *Science* 259:1691-1692, 1993. Techniques for incorporating DNA into such vectors are well known to those of ordinary skill in the art. A retroviral vector may additionally transfer or incorporate a gene for a selectable marker (to aid in the identification or selection of transduced cells) and/or a targeting moiety, such as a gene that encodes a ligand for a receptor on a specific target cell, to render the vector target specific. Targeting may also be accomplished using an antibody, by methods known to those of ordinary skill in the art.

Other formulations for therapeutic purposes include colloidal dispersion systems, such as macromolecule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, and liposomes. A preferred colloidal system for use as a delivery vehicle *in vitro* and *in vivo* is a liposome (*i.e.*, an artificial membrane vesicle). The uptake of naked polynucleotides may be increased by incorporating the polynucleotides into and/or onto biodegradable beads, which are efficiently transported into the cells. The preparation and use of such systems is well known in the art.

In a related aspect, a polynucleotide vaccine as described above may be administered simultaneously with or sequentially to either a polypeptide of the present invention or a known *Chlamydia* antigen. For example, administration of polynucleotides encoding a polypeptide of the present invention, either "naked" or in a delivery system as described above, may be followed by administration of an antigen in order to enhance the protective immune effect of the vaccine.

Polypeptides and polynucleotides disclosed herein may also be employed in adoptive immunotherapy for the treatment of *Chlamydial* infection. Adoptive immunotherapy may be broadly classified into either active or passive immunotherapy. In active immunotherapy, treatment relies on the *in vivo* stimulation of the endogenous host immune system with the administration of immune response-modifying agents (for example, vaccines, bacterial adjuvants, and/or cytokines).

In passive immunotherapy, treatment involves the delivery of biologic reagents with established immune reactivity (such as effector cells or antibodies) that can directly or indirectly mediate anti-*Chlamydia* effects and does not necessarily depend on an intact host immune system. Examples of effector cells include T lymphocytes (for example, CD8+ cytotoxic T-lymphocyte, CD4+ T-helper), killer cells (such as Natural Killer cells, lymphokine-activated killer cells), B cells, or antigen presenting cells (such as dendritic cells and macrophages) expressing the disclosed antigens. The polypeptides disclosed herein may also be used to generate antibodies or anti-idiotypic antibodies (as in U.S. Patent No. 4,918,164), for passive immunotherapy.

The predominant method of procuring adequate numbers of T-cells for adoptive immunotherapy is to grow immune T-cells *in vitro*. Culture conditions for expanding single antigen-specific T-cells to several billion in number with retention of antigen recognition *in vivo* are well known in the art. These *in vitro* culture conditions typically utilize intermittent stimulation with antigen, often in the presence of cytokines, such as IL-2, and non-dividing feeder cells. As noted above, the immunoreactive polypeptides described herein may be used to rapidly expand antigen-specific T cell cultures in order to generate sufficient number of cells for immunotherapy. In particular, antigen-presenting cells, such as dendritic, macrophage, monocyte, fibroblast, or B-cells, may be pulsed with immunoreactive polypeptides, or polynucleotide sequence(s) may be

introduced into antigen presenting cells, using a variety of standard techniques well known in the art. For example, antigen presenting cells may be transfected or transduced with a polynucleotide sequence, wherein said sequence contains a promoter region appropriate for increasing expression, and can be expressed as part of a recombinant virus or other expression system. Several viral vectors may be used to transduce an antigen presenting cell, including pox virus, vaccinia virus, and adenovirus; also, antigen presenting cells may be transfected with polynucleotide sequences disclosed herein by a variety of means, including gene-gun technology, lipid-mediated delivery, electroporation, osmotic shock, and particulate delivery mechanisms, resulting in efficient and acceptable expression levels as determined by one of ordinary skill in the art. For cultured T-cells to be effective in therapy, the cultured T-cells must be able to grow and distribute widely and to survive long term *in vivo*. Studies have demonstrated that cultured T-cells can be induced to grow *in vivo* and to survive long term in substantial numbers by repeated stimulation with antigen supplemented with IL-2 (see, for example, Cheever, M., *et al*, "Therapy With Cultured T Cells: Principles Revisited," *Immunological Reviews*, 157:177, 1997).

The polypeptides disclosed herein may also be employed to generate and/or isolate chlamydial-reactive T-cells, which can then be administered to the patient. In one technique, antigen-specific T-cell lines may be generated by *in vivo* immunization with short peptides corresponding to immunogenic portions of the disclosed polypeptides. The resulting antigen specific CD8+ or CD4+ T-cell clones may be isolated from the patient, expanded using standard tissue culture techniques, and returned to the patient.

Alternatively, peptides corresponding to immunogenic portions of the polypeptides may be employed to generate *Chlamydia* reactive T cell subsets by selective *in vitro* stimulation and expansion of autologous T cells to provide antigen-specific T cells which may be subsequently transferred to the patient as described, for example, by Chang *et al*, (*Crit. Rev. Oncol. Hematol.*, 22(3), 213, 1996). Cells of the immune system, such as T cells, may be isolated from the peripheral blood of a patient, using a commercially available cell separation system, such as Isolex™ System, available from Nexell Therapeutics, Inc. Irvine, CA. The separated cells are stimulated with one or more of the immunoreactive polypeptides contained within a delivery vehicle, such as a microsphere, to provide antigen-specific T cells. The population of antigen-specific T cells is then

expanded using standard techniques and the cells are administered back to the patient.

In other embodiments, T-cell and/or antibody receptors specific for the polypeptides disclosed herein can be cloned, expanded, and transferred into other vectors or effector cells for use in adoptive immunotherapy. In particular, T cells may be transfected  
 5 with the appropriate genes to express the variable domains from chlamydia specific monoclonal antibodies as the extracellular recognition elements and joined to the T cell receptor signaling chains, resulting in T cell activation, specific lysis, and cytokine release. This enables the T cell to redirect its specificity in an MHC-independent manner. See for example, Eshhar, Z., *Cancer Immunol Immunother*, 45(3-4):131-6, 1997 and Hwu, P., et al,  
 10 *Cancer Res*, 55(15):3369-73, 1995. Another embodiment may include the transfection of chlamydia antigen specific alpha and beta T cell receptor chains into alternate T cells, as in Cole, DJ, et al, *Cancer Res*, 55(4):748-52, 1995.

In a further embodiment, syngeneic or autologous dendritic cells may be pulsed with peptides corresponding to at least an immunogenic portion of a polypeptide  
 15 disclosed herein. The resulting antigen-specific dendritic cells may either be transferred into a patient, or employed to stimulate T cells to provide antigen-specific T cells which may, in turn, be administered to a patient. The use of peptide-pulsed dendritic cells to generate antigen-specific T cells and the subsequent use of such antigen-specific T cells to eradicate disease in a murine model has been demonstrated by Cheever et al,  
 20 *Immunological Reviews*, 157:177, 1997). Additionally, vectors expressing the disclosed polynucleotides may be introduced into stem cells taken from the patient and clonally propagated *in vitro* for autologous transplant back into the same patient.

Within certain aspects, polypeptides, polynucleotides, T cells and/or binding agents disclosed herein may be incorporated into pharmaceutical compositions or  
 25 immunogenic compositions (*i.e.*, vaccines). Alternatively, a pharmaceutical composition may comprise an antigen-presenting cell (*e.g.* a dendritic cell) transfected with a *Chlamydial* polynucleotide such that the antigen presenting cell expresses a *Chlamydial* polypeptide. Pharmaceutical compositions comprise one or more such compounds and a physiologically acceptable carrier. Vaccines may comprise one or more such compounds  
 30 and an immunostimulant. An immunostimulant may be any substance that enhances or potentiates an immune response to an exogenous antigen. Examples of immunostimulants

include adjuvants, biodegradable microspheres (e.g., polylactic galactide) and liposomes (into which the compound is incorporated; *see e.g.*, Fullerton, U.S. Patent No. 4,235,877). Vaccine preparation is generally described in, for example, M.F. Powell and M.J. Newman, eds., "Vaccine Design (the subunit and adjuvant approach)," Plenum Press (NY, 1995).

5 Pharmaceutical compositions and vaccines within the scope of the present invention may also contain other compounds, which may be biologically active or inactive. For example, one or more immunogenic portions of other *Chlamydial* antigens may be present, either incorporated into a fusion polypeptide or as a separate compound, within the composition or vaccine.

10 A pharmaceutical composition or vaccine may contain DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated *in situ*. As noted above, the DNA may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacteria and viral expression systems. Numerous gene delivery techniques are well known  
15 in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems* 15:143-198, 1998, and references cited therein. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an  
20 immunogenic portion of the polypeptide on its cell surface or secretes such an epitope.

In a preferred embodiment, the DNA may be introduced using a viral expression system (e.g., vaccinia or other pox virus, retrovirus, adenovirus, baculovirus, togavirus, bacteriophage, and the like), which often involves the use of a non-pathogenic (defective), replication competent virus.

25 For example, many viral expression vectors are derived from viruses of the retroviridae family. This family includes the murine leukemia viruses, the mouse mammary tumor viruses, the human foamy viruses, Rous sarcoma virus, and the immunodeficiency viruses, including human, simian, and feline. Considerations when designing retroviral expression vectors are discussed in Comstock *et al.* (1997).

30 Excellent murine leukemia virus (MLV)-based viral expression vectors have been developed by Kim *et al.* (1998). In creating the MLV vectors, Kim *et al.* found that

the entire *gag* sequence, together with the immediate upstream region, could be deleted without significantly affecting viral packaging or gene expression. Further, it was found that nearly the entire U3 region could be replaced with the immediately-early promoter of human cytomegalovirus without deleterious effects. Additionally, MCR and internal  
5 ribosome entry sites (IRES) could be added without adverse effects. Based on their observations, Kim *et al.* have designed a series of MLV-based expression vectors comprising one or more of the features described above.

As more has been learned about human foamy virus (HFV), characteristics of HFV that are favorable for its use as an expression vector have been discovered. These  
10 characteristics include the expression of pol by splicing and start of translation at a defined initiation codon. Other aspects of HFV viral expression vectors are reviewed in Bodem *et al.* (1997).

Murakami *et al.* (1997) describe a Rous sarcoma virus (RSV)-based replication-competent avian retrovirus vectors, IR1 and IR2 to express a heterologous gene  
15 at a high level. In these vectors, the IRES derived from encephalomyocarditis virus (EMCV) was inserted between the *env* gene and the heterologous gene. The IR1 vector retains the splice-acceptor site that is present downstream of the *env* gene while the IR2 vector lacks it. Murakami *et al.* have shown high level expression of several different heterologous genes by these vectors.

Recently, a number of lentivirus-based retroviral expression vectors have been developed. Kafri *et al.* (1997) have shown sustained expression of genes delivered directly into liver and muscle by a human immunodeficiency virus (HIV)-based expression vector. One benefit of the system is the inherent ability of HIV to transduce non-dividing cells. Because the viruses of Kafri *et al.* are pseudotyped with vesicular stomatitis virus G  
25 glycoprotein (VSVG), they can transduce a broad range of tissues and cell types.

A large number of adenovirus-based expression vectors have been developed, primarily due to the advantages offered by these vectors in gene therapy applications. Adenovirus expression vectors and methods of using such vectors are the subject of a number of United States patents, including United States Patent No. 5,698,202,  
30 United States Patent No. 5,616,326, United States Patent No. 5,585,362, and United States Patent No. 5,518,913, all incorporated herein by reference.

Additional adenoviral constructs are described in Khatri *et al.* (1997) and Tomanin *et al.* (1997). Khatri *et al.* describe novel ovine adenovirus expression vectors and their ability to infect bovine nasal turbinate and rabbit kidney cells as well as a range of human cell type, including lung and foreskin fibroblasts as well as liver, prostate, breast, colon and retinal lines. Tomanin *et al.* describe adenoviral expression vectors containing the T7 RNA polymerase gene. When introduced into cells containing a heterologous gene operably linked to a T7 promoter, the vectors were able to drive gene expression from the T7 promoter. The authors suggest that this system may be useful for the cloning and expression of genes encoding cytotoxic proteins.

Poxviruses are widely used for the expression of heterologous genes in mammalian cells. Over the years, the vectors have been improved to allow high expression of the heterologous gene and simplify the integration of multiple heterologous genes into a single molecule. In an effort to diminish cytopathic effects and to increase safety, vaccinia virus mutant and other poxviruses that undergo abortive infection in mammalian cells are receiving special attention (Oertli *et al.*, 1997). The use of poxviruses as expression vectors is reviewed in Carroll and Moss (1997).

Togaviral expression vectors, which includes alphaviral expression vectors have been used to study the structure and function of proteins and for protein production purposes. Attractive features of togaviral expression vectors are rapid and efficient gene expression, wide host range, and RNA genomes (Huang, 1996). Also, recombinant vaccines based on alphaviral expression vectors have been shown to induce a strong humoral and cellular immune response with good immunological memory and protective effects (Tubulekas *et al.*, 1997). Alphaviral expression vectors and their use are discussed, for example, in Lundstrom (1997).

In one study, Li and Garoff (1996) used Semliki Forest virus (SFV) expression vectors to express retroviral genes and to produce retroviral particles in BHK-21 cells. The particles produced by this method had protease and reverse transcriptase activity and were infectious. Furthermore, no helper virus could be detected in the virus stocks. Therefore, this system has features that are attractive for its use in gene therapy protocols.

Baculoviral expression vectors have traditionally been used to express heterologous proteins in insect cells. Examples of proteins include mammalian chemokine



receptors (Wang *et al.*, 1997), reporter proteins such as green fluorescent protein (Wu *et al.*, 1997), and FLAG fusion proteins (Wu *et al.*, 1997; Koh *et al.*, 1997). Recent advances in baculoviral expression vector technology, including their use in virion display vectors and expression in mammalian cells is reviewed by Possee (1997). Other reviews on baculoviral  
 5 expression vectors include Jones and Morikawa (1996) and O'Reilly (1997).

Other suitable viral expression systems are disclosed, for example, in Fisher-Hoch *et al.*, *Proc. Natl. Acad. Sci. USA* 86:317-321, 1989; Flexner *et al.*, *Ann. N.Y. Acad. Sci.* 569:86-103, 1989; Flexner *et al.*, *Vaccine* 8:17-21, 1990; U.S. Patent Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127;  
 10 GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, *Biotechniques* 6:616-627, 1988; Rosenfeld *et al.*, *Science* 252:431-434, 1991; Kolls *et al.*, *Proc. Natl. Acad. Sci. USA* 91:215-219, 1994; Kass-Eisler *et al.*, *Proc. Natl. Acad. Sci. USA* 90:11498-11502, 1993; Guzman *et al.*, *Circulation* 88:2838-2848, 1993; and Guzman *et al.*, *Cir. Res.* 73:1202-1207, 1993. Techniques for incorporating DNA into such expression systems are  
 15 well known to those of ordinary skill in the art. In other systems, the DNA may be introduced as "naked" DNA, as described, for example, in Ulmer *et al.*, *Science* 259:1745-1749, 1993 and reviewed by Cohen, *Science* 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

20 It will be apparent that a vaccine may comprise a polynucleotide and/or a polypeptide component, as desired. It will also be apparent that a vaccine may contain pharmaceutically acceptable salts of the polynucleotides and/or polypeptides provided herein. Such salts may be prepared from pharmaceutically acceptable non-toxic bases, including organic bases (*e.g.*, salts of primary, secondary and tertiary amines and basic  
 25 amino acids) and inorganic bases (*e.g.*, sodium, potassium, lithium, ammonium, calcium and magnesium salts). While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will vary depending on the mode of administration. Compositions of the present invention may be formulated for any appropriate manner of administration, including for example,  
 30 topical, oral, nasal, intravenous, intracranial, intraperitoneal, subcutaneous or intramuscular administration. For parenteral administration, such as subcutaneous injection, the carrier

preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier, such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (*e.g.*, polylactate polyglycolate) may also be employed as carriers for the pharmaceutical compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268 and 5,075,109.

Such compositions may also comprise buffers (*e.g.*, neutral buffered saline or phosphate buffered saline), carbohydrates (*e.g.*, glucose, mannose, sucrose or dextrans), mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, bacteriostats, chelating agents such as EDTA or glutathione, adjuvants (*e.g.*, aluminum hydroxide), solutes that render the formulation isotonic, hypotonic or weakly hypertonic with the blood of a recipient, suspending agents, thickening agents and/or preservatives. Alternatively, compositions of the present invention may be formulated as a lyophilizate. Compounds may also be encapsulated within liposomes using well known technology.

Any of a variety of immunostimulants may be employed in the vaccines of this invention. For example, an adjuvant may be included. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bordetella pertussis* or *Mycobacterium tuberculosis* derived proteins. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham, Philadelphia, PA); aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF or interleukin-2, -7, or -12, may also be used as adjuvants.

Within the vaccines provided herein, under select circumstances, the adjuvant composition may be designed to induce an immune response predominantly of the Th1 type or Th2 type. High levels of Th1-type cytokines (*e.g.*, IFN- $\gamma$ , TNF $\alpha$ , IL-2 and IL-

12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (e.g., IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines will increase to a greater extent than the level of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann and Coffman, *Ann. Rev. Immunol.* 7:145-173, 1989.

Preferred adjuvants for use in eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A (3D-MPL), together with an aluminum salt. MPL adjuvants are available from Corixa Corporation (Seattle, WA; see US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, for example, in WO 96/02555 and WO 99/33488. Immunostimulatory DNA sequences are also described, for example, by Sato et al., *Science* 273:352, 1996. Another preferred adjuvant is a saponin, preferably QS21 (Aquila Biopharmaceuticals Inc., Framingham, MA), which may be used alone or in combination with other adjuvants. For example, an enhanced system involves the combination of a monophosphoryl lipid A and saponin derivative, such as the combination of QS21 and 3D-MPL as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. A particularly potent adjuvant formulation involving QS21, 3D-MPL and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Other preferred adjuvants include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (e.g., SBAS-2 or SBAS-4, available from SmithKline Beecham, Rixensart, Belgium), Detox (Corixa Corporation; Seattle, WA), RC-529 (Corixa Corporation; Seattle, WA) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such

as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are incorporated herein by reference in their entireties.

Any vaccine provided herein may be prepared using well known methods that result in a combination of antigen, immunostimulant and a suitable carrier or excipient.

5 The compositions described herein may be administered as part of a sustained release formulation (*i.e.*, a formulation such as a capsule, sponge or gel (composed of polysaccharides, for example) that effects a slow release of compound following administration). Such formulations may generally be prepared using well known technology (*see, e.g.*, Coombes et al., *Vaccine* 14:1429-1438, 1996) and administered by,  
10 for example, oral, rectal or subcutaneous implantation, or by implantation at the desired target site. Sustained-release formulations may contain a polypeptide, polynucleotide or antibody dispersed in a carrier matrix and/or contained within a reservoir surrounded by a rate controlling membrane.

Carriers for use within such formulations are biocompatible, and may also  
15 be biodegradable; preferably the formulation provides a relatively constant level of active component release. Such carriers include microparticles of poly(lactide-co-glycolide), as well as polyacrylate, latex, starch, cellulose and dextran. Other delayed-release carriers include supramolecular biovectors, which comprise a non-liquid hydrophilic core (*e.g.*, a cross-linked polysaccharide or oligosaccharide) and, optionally, an external layer  
20 comprising an amphiphilic compound, such as a phospholipid (*see e.g.*, U.S. Patent No. 5,151,254 and PCT applications WO 94/20078, WO/94/23701 and WO 96/06638). The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of the condition to be treated or prevented.

25 Any of a variety of delivery vehicles may be employed within pharmaceutical compositions and vaccines to facilitate production of an antigen-specific immune response that targets *Chlamydia*-infected cells. Delivery vehicles include antigen presenting cells (APCs), such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be  
30 genetically modified to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-*Chlamydia* effects *per se*

and/or to be immunologically compatible with the receiver (*i.e.*, matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or  
 5 progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau and Steinman, *Nature* 392:245-251, 1998) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic immunity (*see* Timmerman and Levy, *Ann. Rev. Med.* 50:507-529, 1999). In general, dendritic cells may be identified based on their typical shape (stellate *in situ*, with marked cytoplasmic  
 10 processes (dendrites) visible *in vitro*), their ability to take up, process and present antigens with high efficiency, and their ability to activate naïve T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As an alternative to dendritic cells, secreted  
 15 vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (*see* Zitvogel et al., *Nature Med.* 4:594-600, 1998).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination  
 20 of cytokines such as GM-CSF, IL-4, IL-13 and/or TNF $\alpha$  to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNF $\alpha$ , CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation  
 25 and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC  
 30 with a high capacity for antigen uptake and processing, which correlates with the high expression of Fc $\gamma$  receptor and mannose receptor. The mature phenotype is typically

characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (*e.g.*, CD54 and CD11) and costimulatory molecules (*e.g.*, CD40, CD80, CD86 and 4-1BB).

- 5           APCs may generally be transfected with a polynucleotide encoding a *Chlamydial* protein (or portion or other variant thereof) such that the *Chlamydial* polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a composition or vaccine comprising such transfected cells may then be used for therapeutic purposes, as described herein.
- 10   Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi et al., *Immunology and cell Biology* 75:456-460, 1997.
- 15   Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with the *Chlamydial* polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or viruses (*e.g.*, vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (*e.g.*, a carrier
- 20   molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

Routes and frequency of administration of pharmaceutical compositions and vaccines, as well as dosage, will vary from individual to individual. In general, the pharmaceutical compositions and vaccines may be administered by injection (*e.g.*,  
 25   intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (*e.g.*, by aspiration) or orally. Between 1 and 3 doses may be administered for a 1-36 week period. Preferably, 3 doses are administered, at intervals of 3-4 months, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of polypeptide or DNA that, when administered as  
 30   described above, is capable of raising an immune response in an immunized patient sufficient to protect the patient from *Chlamydial* infection for at least 1-2 years. In general,

the amount of polypeptide present in a dose (or produced *in situ* by the DNA in a dose) ranges from about 1 pg to about 100 mg per kg of host, typically from about 10 pg to about 1 mg, and preferably from about 100 pg to about 1 µg. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

5 While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will vary depending on the mode of administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier, such as  
10 mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (*e.g.*, polylactic galactide) may also be employed as carriers for the pharmaceutical compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268 and 5,075,109.

15 In general, an appropriate dosage and treatment regimen provides the active compound(s) in an amount sufficient to provide therapeutic and/or prophylactic benefit. Such a response can be monitored by establishing an improved clinical outcome in treated patients as compared to non-treated patients. Increases in preexisting immune responses to a *Chlamydial* protein generally correlate with an improved clinical outcome. Such immune  
20 responses may generally be evaluated using standard proliferation, cytotoxicity or cytokine assays, which may be performed using samples obtained from a patient before and after treatment.

#### DETECTION AND DIAGNOSIS

25 In another aspect, the present invention provides methods for using the polypeptides described above to diagnose Chlamydial infection. In this aspect, methods are provided for detecting Chlamydial infection in a biological sample, using one or more of the above polypeptides, either alone or in combination. For clarity, the term "polypeptide" will be used when describing specific embodiments of the inventive  
30 diagnostic methods. However, it will be clear to one of skill in the art that the fusion proteins of the present invention may also be employed in such methods.

As used herein, a "biological sample" is any antibody-containing sample obtained from a patient. Preferably, the sample is whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid or urine. More preferably, the sample is a blood, serum or plasma sample obtained from a patient. The polypeptides are used in an assay, as described  
5 below, to determine the presence or absence of antibodies to the polypeptide(s) in the sample, relative to a predetermined cut-off value. The presence of such antibodies indicates previous sensitization to *Chlamydia* antigens which may be indicative of *Chlamydia*-infection.

In embodiments in which more than one polypeptide is employed, the  
10 polypeptides used are preferably complementary (*i.e.*, one component polypeptide will tend to detect infection in samples where the infection would not be detected by another component polypeptide). Complementary polypeptides may generally be identified by using each polypeptide individually to evaluate serum samples obtained from a series of patients known to be infected with *Chlamydia*. After determining which samples test  
15 positive (as described below) with each polypeptide, combinations of two or more polypeptides may be formulated that are capable of detecting infection in most, or all, of the samples tested.

A variety of assay formats are known to those of ordinary skill in the art for using one or more polypeptides to detect antibodies in a sample. *See, e.g.*, Harlow and  
20 Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988, which is incorporated herein by reference. In a preferred embodiment, the assay involves the use of polypeptide immobilized on a solid support to bind to and remove the antibody from the sample. The bound antibody may then be detected using a detection reagent that contains a reporter group. Suitable detection reagents include antibodies that bind to the  
25 antibody/polypeptide complex and free polypeptide labeled with a reporter group (*e.g.*, in a semi-competitive assay). Alternatively, a competitive assay may be utilized, in which an antibody that binds to the polypeptide is labeled with a reporter group and allowed to bind to the immobilized antigen after incubation of the antigen with the sample. The extent to which components of the sample inhibit the binding of the labeled antibody to the  
30 polypeptide is indicative of the reactivity of the sample with the immobilized polypeptide.



The solid support may be any solid material known to those of ordinary skill in the art to which the antigen may be attached. For example, the solid support may be a test well in a microtiter plate, or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681.

The polypeptides may be bound to the solid support using a variety of techniques known to those of ordinary skill in the art. In the context of the present invention, the term "bound" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the antigen and functional groups on the support or may be a linkage by way of a cross-linking agent). Binding by adsorption to a well in a microtiter plate or to a membrane is preferred. In such cases, adsorption may be achieved by contacting the polypeptide, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of polypeptide ranging from about 10 ng to about 1  $\mu$ g, and preferably about 100 ng, is sufficient to bind an adequate amount of antigen.

Covalent attachment of polypeptide to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the polypeptide. For example, the polypeptide may be bound to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the polypeptide (*see, e.g.*, Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is an enzyme linked immunosorbent assay (ELISA). This assay may be performed by first contacting a polypeptide antigen that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that antibodies to the polypeptide within the sample are allowed to bind to the immobilized polypeptide. Unbound sample is then removed from the immobilized polypeptide and a detection reagent capable of binding to the immobilized antibody-

polypeptide complex is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific detection reagent.

More specifically, once the polypeptide is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin (BSA) or Tween 20™ (Sigma Chemical Co., St. Louis, MO) may be employed. The immobilized polypeptide is then incubated with the sample, and antibody is allowed to bind to the antigen. The sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact time (*i.e.*, incubation time) is that period of time that is sufficient to detect the presence of antibody within an HGE-infected sample. Preferably, the contact time is sufficient to achieve a level of binding that is at least 95% of that achieved at equilibrium between bound and unbound antibody. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. Detection reagent may then be added to the solid support. An appropriate detection reagent is any compound that binds to the immobilized antibody-polypeptide complex and that can be detected by any of a variety of means known to those in the art. Preferably, the detection reagent contains a binding agent (such as, for example, Protein A, Protein G, immunoglobulin, lectin or free antigen) conjugated to a reporter group. Preferred reporter groups include enzymes (such as horseradish peroxidase), substrates, cofactors, inhibitors, dyes, radionuclides, luminescent groups, fluorescent groups and biotin. The conjugation of binding agent to reporter group may be achieved using standard methods known to those of ordinary skill in the art. Common binding agents may also be purchased conjugated to a variety of reporter groups from many commercial sources (*e.g.*, Zymed Laboratories, San Francisco, CA, and Pierce, Rockford, IL).

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound antibody. An appropriate amount of time may generally be determined from the manufacturer's instructions or by assaying the level of binding that occurs over a period of time. Unbound  
 5 detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups and fluorescent groups. Biotin may be detected using  
 10 avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of anti-*Chlamydia* antibodies in the  
 15 sample, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value is the average mean signal obtained when the immobilized antigen is incubated with samples from an uninfected patient. In general, a sample generating a signal that is three standard deviations above the predetermined cut-off  
 20 value is considered positive for *Chlamydia*-infection. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, pp. 106-107. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and  
 25 false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along  
 30 the plot, to minimize the false positive rate, or to the right, to minimize the false negative

rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for Chlamydial infection.

In a related embodiment, the assay is performed in a rapid flow-through or strip test format, wherein the antigen is immobilized on a membrane, such as nitrocellulose.

5 In the flow-through test, antibodies within the sample bind to the immobilized polypeptide as the sample passes through the membrane. A detection reagent (*e.g.*, protein A-colloidal gold) then binds to the antibody-polypeptide complex as the solution containing the detection reagent flows through the membrane. The detection of bound detection reagent may then be performed as described above. In the strip test format, one end of the

10 membrane to which polypeptide is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing detection reagent and to the area of immobilized polypeptide. Concentration of detection reagent at the polypeptide indicates the presence of anti-*Chlamydia* antibodies in the sample. Typically, the concentration of detection reagent at that site generates a pattern, such as a line, that can

15 be read visually. The absence of such a pattern indicates a negative result. In general, the amount of polypeptide immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of antibodies that would be sufficient to generate a positive signal in an ELISA, as discussed above. Preferably, the amount of polypeptide immobilized on the membrane ranges from about 25 ng to about 1  $\mu$

20 g, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount (*e.g.*, one drop) of patient serum or blood.

Of course, numerous other assay protocols exist that are suitable for use with the polypeptides of the present invention. The above descriptions are intended to be exemplary only. One example of an alternative assay protocol which may be usefully

25 employed in such methods is a Western blot, wherein the proteins present in a biological sample are separated on a gel, prior to exposure to a binding agent. Such techniques are well known to those of skill in the art.

#### BINDING AGENTS AND THEIR USES

30 The present invention further provides agents, such as antibodies and antigen-binding fragments thereof, that specifically bind to a *Chlamydial* protein. As used

herein, an antibody, or antigen-binding fragment thereof, is said to "specifically bind" to a *Chlamydial* protein if it reacts at a detectable level (within, for example, an ELISA) with a *Chlamydial* protein, and does not react detectably with unrelated proteins under similar conditions. As used herein, "binding" refers to a noncovalent association between two  
5 separate molecules such that a complex is formed. The ability to bind may be evaluated by, for example, determining a binding constant for the formation of the complex. The binding constant is the value obtained when the concentration of the complex is divided by the product of the component concentrations. In general, two compounds are said to "bind," in the context of the present invention, when the binding constant for complex formation  
10 exceeds about  $10^3$  L/mol. The binding constant may be determined using methods well known in the art.

Binding agents may be further capable of differentiating between patients with and without a *Chlamydial* infection using the representative assays provided herein. In other words, antibodies or other binding agents that bind to a *Chlamydial* protein will  
15 generate a signal indicating the presence of a *Chlamydial* infection in at least about 20% of patients with the disease, and will generate a negative signal indicating the absence of the disease in at least about 90% of individuals without infection. To determine whether a binding agent satisfies this requirement, biological samples (*e.g.*, blood, sera, sputum urine and/or tissue biopsies ) from patients with and without *Chlamydial* infection (as determined  
20 using standard clinical tests) may be assayed as described herein for the presence of polypeptides that bind to the binding agent. It will be apparent that a statistically significant number of samples with and without the disease should be assayed. Each binding agent should satisfy the above criteria; however, those of ordinary skill in the art will recognize that binding agents may be used in combination to improve sensitivity.

25 Any agent that satisfies the above requirements may be a binding agent. For example, a binding agent may be a ribosome, with or without a peptide component, an RNA molecule or a polypeptide. In a preferred embodiment, a binding agent is an antibody or an antigen-binding fragment thereof. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. *See, e.g.*, Harlow and Lane,  
30 *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, antibodies can be produced by cell culture techniques, including the generation of

monoclonal antibodies as described herein, or via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies. In one technique, an immunogen comprising the polypeptide is initially injected into any of a wide variety of mammals (e.g., mice, rats, rabbits, sheep or goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for an antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976, and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the desired specificity (i.e., reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and their culture supernatants tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the

ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

5           Within certain embodiments, the use of antigen-binding fragments of antibodies may be preferred. Such fragments include Fab fragments, which may be prepared using standard techniques. Briefly, immunoglobulins may be purified from rabbit serum by affinity chromatography on Protein A bead columns (Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988) and digested by  
10       papain to yield Fab and Fc fragments. The Fab and Fc fragments may be separated by affinity chromatography on protein A bead columns.

          Monoclonal antibodies of the present invention may be coupled to one or more therapeutic agents. Suitable agents in this regard include radionuclides, differentiation inducers, drugs, toxins, and derivatives thereof. Preferred radionuclides  
15       include  $^{90}\text{Y}$ ,  $^{123}\text{I}$ ,  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{186}\text{Re}$ ,  $^{188}\text{Re}$ ,  $^{211}\text{At}$ , and  $^{212}\text{Bi}$ . Preferred drugs include methotrexate, and pyrimidine and purine analogs. Preferred differentiation inducers include phorbol esters and butyric acid. Preferred toxins include ricin, abrin, diphtheria toxin, cholera toxin, gelonin, *Pseudomonas* exotoxin, *Shigella* toxin, and pokeweed antiviral protein.

20           A therapeutic agent may be coupled (*e.g.*, covalently bonded) to a suitable monoclonal antibody either directly or indirectly (*e.g.*, via a linker group). A direct reaction between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one may be capable of reacting with a carbonyl-containing group,  
25       such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (*e.g.*, a halide) on the other.

          Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can  
30       also serve to increase the chemical reactivity of a substituent on an agent or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also

facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the catalog of the Pierce Chemical Co., Rockford, IL), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, *e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.

Where a therapeutic agent is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (*e.g.*, U.S. Patent No. 4,489,710, to Spitler), by irradiation of a photolabile bond (*e.g.*, U.S. Patent No. 4,625,014, to Senter et al.), by hydrolysis of derivatized amino acid side chains (*e.g.*, U.S. Patent No. 4,638,045, to Kohn et al.), by serum complement-mediated hydrolysis (*e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.), and acid-catalyzed hydrolysis (*e.g.*, U.S. Patent No. 4,569,789, to Blattler et al.).

It may be desirable to couple more than one agent to an antibody. In one embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody. Regardless of the particular embodiment, immunoconjugates with more than one agent may be prepared in a variety of ways. For example, more than one agent may be coupled directly to an antibody molecule, or linkers which provide multiple sites for attachment can be used. Alternatively, a carrier can be used.

A carrier may bear the agents in a variety of ways, including covalent bonding either directly or via a linker group. Suitable carriers include proteins such as albumins (*e.g.*, U.S. Patent No. 4,507,234, to Kato et al.), peptides and polysaccharides such as aminodextran (*e.g.*, U.S. Patent No. 4,699,784, to Shih et al.). A carrier may also bear an agent by noncovalent bonding or by encapsulation, such as within a liposome vesicle (*e.g.*, U.S. Patent Nos. 4,429,008 and 4,873,088). Carriers specific for radionuclide



agents include radiohalogenated small molecules and chelating compounds. For example, U.S. Patent No. 4,735,792 discloses representative radiohalogenated small molecules and their synthesis. A radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, or metal oxide, radionuclide. For example, U.S. Patent No. 4,673,562, to Davison et al. discloses representative chelating compounds and their synthesis.

A variety of routes of administration for the antibodies and immunoconjugates may be used. Typically, administration will be intravenous, intramuscular, subcutaneous or in site-specific regions by appropriate methods. It will be evident that the precise dose of the antibody/immunoconjugate will vary depending upon the antibody used, the antigen density, and the rate of clearance of the antibody.

Antibodies may be used in diagnostic tests to detect the presence of *Chlamydia* antigens using assays similar to those detailed above and other techniques well known to those of skill in the art, thereby providing a method for detecting Chlamydial infection in a patient.

Diagnostic reagents of the present invention may also comprise DNA sequences encoding one or more of the above polypeptides, or one or more portions thereof. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify *Chlamydia*-specific cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for a DNA molecule encoding a polypeptide of the present invention. The presence of the amplified cDNA is then detected using techniques well known in the art, such as gel electrophoresis. Similarly, oligonucleotide probes specific for a DNA molecule encoding a polypeptide of the present invention may be used in a hybridization assay to detect the presence of an inventive polypeptide in a biological sample.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLE 1CD4 T CELL EXPRESSION CLONING FOR THE IDENTIFICATION OF T CELL STIMULATING  
ANTIGENS FROM CHLAMYDIA TRACHOMATIS SEROVAR E

5 In this example, a CD4+ T cell expression cloning strategy was used to identify Chlamydia trachomatis antigens recognized by patients enrolled in Corixa Corporation's blood donor program. A genomic library of Chlamydia trachomatis serovar E was constructed and screened with Chlamydia specific T cell lines generated by stimulating PBMCs from these donors. Donor CT1 is a 27 yr. old male whose clinical  
10 manifestation was non-gonococcal urethritis and his urine was tested positive for Chlamydia by ligase chain reaction. Donor CT3 is a 43 yr. old male who is asymptomatic and infected with serovar J. Donor CT10 is a 24 yr. old female who is asymptomatic and was exposed to Chlamydia through her partner but did not develop the disease. Donor CT11 is a 24 yr. old female with multiple infections (serovar J, F and E).

15 Chlamydia specific T-cell lines were generated from donors with Chlamydia genital tract infection or donors exposed to chlamydia who did not develop the disease. T cell lines from donor CT-1, CT-3 and CT-10 were generated by stimulating PBMCs with reticulate bodies of C. trachomatis serovar E. T-cell lines from donor CT-11 were generated by stimulating PBMCs with either reticulate bodies or elementary bodies of C.  
20 trachomatis serovar E. A randomly sheared genomic library of C. trachomatis serovar E was constructed in lambda Zap II vector and an amplified library plated out in 96 well microtiter plates at a density of 25 clones/well. Bacteria were induced to express the recombinant protein in the presence of 2mM IPTG for 2hr, then pelleted and resuspended in 200ul RPMI/10% FBS. 10 ul of the induced bacterial suspension was transferred to 96  
25 well plates containing autologous monocyte-derived dendritic cells. After a 2 hour incubation, dendritic cells were washed to remove E. coli and the T cells were added. Positive E. coli pools were identified by determining IFN gamma production and proliferation of T cells in the pools. The number of pools identified by each T-cell line is as follows: CT1 line : 30/480 pools; CT3 line : 91/960 pools; CT10 line : 40/480 pools; CT11  
30 line : 51/480 pools. The clones identified using this approach are set forth in SEQ ID NO:1-14.

In another example using substantially the same approach described above, we identified 12 additional T-cell reactive clones from *Chlamydia trachomatis* serovar E expression screening. Clone E5-E9-3 (CT1 positive) contains a 636 bp insert that encodes partially the ORF for dnaK like gene. Part of this sequence was also identified in clone E1-A5-53. Clone E4-H3-56 (CT1 positive, 463 bp insert) contains a partial ORF for the TSA gene (CT603) on the complementary strand. The insert for clone E2-G12-52 (1265 bp) was identified with the CT11 line. It contains a partial ORF for clpB, a protease ATPase. Another clone identified with the CT11 line, E1-F9-79 (167 bp), contains a partial ORF for the gene CT133 on the complementary strand. CT133 is a predicted rRNA methylase. Clone E4-D2-79 (CT3 positive) contains a 1181 bp insert that is a partial ORF for nrdA gene. The ORF for this gene was also identified in clone E2-B10-52 (CT10 positive). Clone E6-C8-95 contains a 731 bp insert that was identified using the donor lines CT3, CT1, and CT12. This insert has a carboxy terminal half for the gene for the 60 kDa ORF. Clone E7-H11-61 (CT3 positive-1135 bp) has partial inserts for fliA (CT061), tyrS (CT062), TSA (CT603) and a hypothetical protein (CT602). The insert for clone E5-A11-8 (CT10 positive-1736 bp) contains the complete ORF for groES (CT111) and a majority of the ORF for groEL (CT110). Clone E3-F2-37 (CT10, CT3, CT11, and CT12 positive-1377bp insert) contains a partial ORF for gene tRNA-Trp (CT322) and a complete ORF for the gene secE (CT321). E4-G9-75 is another CT10 clone that contains a partial ORF (723 bp insert) for the amino terminal region of the pmpH gene (CT872). Clone E2-D5-89 (516bp) is also a CT10 positive clone that contains a partial ORF for pmpD gene (12). The insert for clone E5-E2-10 (CT10 positive) is 427 bp and contains a partial ORF for the major outer membrane protein omp1.

25

EXAMPLE 2ADDITIONAL CD4 T CELL EXPRESSION CLONING FOR THE IDENTIFICATION OF T CELL  
STIMULATING ANTIGENS FROM CHLAMYDIA TRACHOMATIS SEROVAR E

5 Twenty sequences were isolated from single clones using a Chlamydia trachomatis serovar E (Ct E) library expression screening method. Descriptions of how the clones and lines were generated are provided in Example 1.

Clone E5-A8-85 (identified using the CT1 patient line) was found to contain a 1433 bp insert. This insert contains a large region of the C-terminal half of the CT875, a  
10 Chlamydia trachomatis hypothetical specific gene that is disclosed in SEQ ID NO:34. Also present in the clone is a partial open reading frame (ORF) of a hypothetical protein CT001 which is on the complementary strand.

The clone E9-G2-93 (identified using the C10 patient line) was shown to contain a 554 bp insert, the sequence of which is disclosed in SEQ ID NO:33. This  
15 sequence encodes a partial ORF for CT178, a hypothetical CT protein.

Clone E7-B1-16 (identified using the patient lines CT10, CT3, CT5, CT11, CT13, and CHH037) has a 2577 bp insert, the sequence of which is disclosed in SEQ ID NO:32. This clone was found to contain three ORFs. The first ORF contains almost the  
20 entire ORF for CT694, a Chlamydia trachomatis (CT) specific hypothetical protein. The second ORF is a full length ORF for CT695, another hypothetical CT protein. The third ORF is the N-terminal portion of CT696.

Clone E9-D5-8 (identified using the patient lines CT10, CT1, CT4, and CT11) contains a 393 bp insert, which is disclosed in SEQ ID NO:31. It was found to encode a partial ORF for CT680, the S2 ribosomal protein.

25 Clone E9-E10-51 (identified using the patient line CT10) contains an 883 bp insert, the sequence of which is disclosed in SEQ ID NO:30. This clone contains two partial ORF. The first of these is for the C-terminal half of CT680, which may show some overlap with the insert present in clone E9-D5-8. The second ORF is the N-terminal partial ORF for CT679, which is the elongation factor TS.

30 Clone E3-B4-18 (identified using the CT1 patient line) contains a 1224 bp insert, the sequence of which is disclosed in SEQ ID NO:29. This clone contains 4 ORFs.

At the N-terminal end of the clone is the complete ORF for CT772, coding for inorganic pyrophosphatase. The second ORF is a small portion of the C-terminal end of CT771, on the complementary frame. The third is a partial ORF of the hypothetical protein, CT191 and the fourth is a partial ORF for CT190, DNA gyrase-B.

5           Clone E10-B2-57 (identified using the CT10 patient line) contains an 822 bp insert, the sequence of which is disclosed in SEQ ID NO:42. This clone contains the complete ORF for CT066, a hypothetical protein, on the complementary strand.

          Clone E3-F3-18 (identified using the CT1 patient line) contains an 1141 bp insert, the sequence of which is disclosed in SEQ ID NO:41. It contains a partial ORF for  
10   pmpG (CT871) in frame with the  $\beta$ -gal gene.

          Clone E4-D6-21 (identified using the CT3 patient line) contains a 1297 bp insert, the sequence of which is disclosed in SEQ ID NO:40. This clone contains a very small portion of xseA (CT329), the entire ORF for tpiS (CT328) on the complementary strand, and a partial amino terminal ORF for trpC (CT327) on the top frame.

15           Clone E1-G9-23 (identified using the CT3 patient line) contains an 1180 bp insert, the sequence of which is disclosed in SEQ ID NO:39. This clone contains almost the entire ORF for glycogen synthase (CT798).

          Clone E3-A3-31 (identified using the CT1 patient line) contains an 1834 bp insert, the sequence of which is disclosed in SEQ ID NO:38. This clone contains a large  
20   region of the hypothetical gene CT622.

          Clone E2-F7-11 (identified using both the CT3 and CT10 patient lines) contains a 2093 bp insert, the sequence of which is disclosed in SEQ ID NO:37. This clone contains a large region of the rpoN gene (CT609) in frame with  $\beta$ -gal and the complete ORF for the hypothetical gene CT610 on the complementary strand. In addition, it also  
25   contains the carboxy-terminal end of CT611, another hypothetical gene.

          Clone E7-H11-10 (identified using the CT3 patient line) contains a 1990 bp insert, the sequence of which is disclosed in SEQ ID NO:36. This clone contains the amino terminal partial ORF for CT610, a complete ORF for CT611, another complete ORF for CT612, and a carboxy-terminal portion of CT613. All of these genes are hypothetical and  
30   all are present on the complementary strand.

Clone E10-C6-45 (identified using the CT3 patient line) contains a 196 bp insert, the sequence of which is disclosed in SEQ ID NO:35. This clone contains a partial ORF for *nrdA* (CT827) in frame with  $\beta$ -gal. This clone contains a relatively small insert and has particular utility in determining the epitope of this gene that contributes to the immunogenicity of Serovar E.

Clone E3-H6-10 (identified using the CT12 patient line) contains a 3734 bp insert, the sequence of which is disclosed in SEQ ID NO:48. This clone contains ORFs for a series of hypothetical proteins. It contains the partial ORFs for CT223 and CT229 and the complete ORFs for CT224, CT225, CT226, CT227, and CT228.

Clone E4-C3-40 (identified using the CT10 patient line) contains a 2044 bp insert, the sequence of which is disclosed in SEQ ID NO:47. This clone contains a partial ORF for *nrdA* (CT827) and the complete ORF for *nrdB* (CT828).

Clone E2-D8-19 (identified using the CT1 patient line) contains a 2010 bp insert, the sequence of which is disclosed in SEQ ID NO:46. This clone contains ORF from the *Chlamydia trachomatis* plasmid as well as containing partial ORFs for ORF3 and ORF6, and complete ORFs for ORF4 and ORF5.

Clone E3-D10-46 (identified using the patient lines CT1, CT3, CT4, CT11, and CT12) contains a 1666 bp insert, the sequence of which is identified in SEQ ID NO: 45. This clone contains a partial ORF for CT770 (fab F), a complete ORF for CT771 (hydrolase/phosphatase homologue), a complete ORF for CT772 (ppa, inorganic phosphatase), and a partial ORF for CT773 (Idh, Leucine dehydrogenase).

Clone E10-H8-1 (identified using both the CT3 and CT10 patient lines) contains an 1862 bp insert, the sequence of which is disclosed in SEQ ID NO:44. It contains the partial ORFs for CT871 (*pmpG*) as well as CT872 (*pmpH*).

Clone E3-F3-7 (identified using the CT1 patient line) contains a 1643 bp insert, the sequence of which is identified in SEQ ID NO:43. It contains the partial ORFs for both CT869 (*pmpE*) and CT870 (*pmpF*).

EXAMPLE 3ADDITIONAL CD4 T CELL EXPRESSION CLONING FOR THE IDENTIFICATION OF T CELL  
STIMULATING ANTIGENS FROM CHLAMYDIA TRACHOMATIS SEROVAR E

5           The T cell line CHH037 was generated from a 22 year-old healthy female sero-negative for Chlamydia. This line was used to screen the Chlamydia trachomatis serovar E library. Nineteen clones were identified from this screen, as described below.

          Clone E7-B12-65, contains an 1179 bp insert, the sequence of which is disclosed in SEQ ID NO:114. It contains the complete ORF of the gene for Malate  
10   dehydrogenase (CT376) on the complementary strand.

          Clone E4-H9-83 contains a 772 bp insert, the sequence of which is identified in SEQ ID NO:115. It contains the partial ORF for the heat shock protein GroEL (CT110).

          Clone E9-B10-52 contains a 487 bp insert, the sequence of which is  
15   identified in SEQ ID NO:116. It contains a partial ORF for the gene yscC (CT674), a general secretion pathway protein.

          Clone E7-A7-79 contains a 1014 bp insert, the sequence of which is disclosed in SEQ ID NO:117. It contains the complete ORF for the histone like development gene, hctA (CT743) and a partial ORF for the rRNA methyltransferase gene  
20   ygcA (CT742).

          Clone E2-D11-18 contains a 287 bp insert, the sequence of which is disclosed in SEQ ID NO:118. It contains the partial ORF for hctA (CT743).

          Clone E9-H6-15, identified using the CT3 line, contains a 713 bp insert the sequence of which is disclosed in SEQ ID NO:125. It contains the partial ORF of the  
25   pmpB gene (CT413).

          Clone E3-D10-87, identified using the CT1 line, contains a 780 bp insert, the sequence of which is disclosed in SEQ ID NO:126. It contains the partial ORF for CT388, a hypothetical gene, on the complementary strand, and a partial ORF for CT389, another hypothetical protein.

30           Clone E9-D6-43, identified using the CT3 line, contains a 433 bp insert, the sequence of which is disclosed in SEQ ID NO:127. It contains a partial ORF for CT858.

Clone E3-D10-4, identified using the CT1 line, contains an 803 bp insert, the sequence of which is disclosed in SEQ ID NO:128. It contains a partial ORF for pGP3-D, an ORF encoded on the plasmid pCHL1.

Clone E3-G8-7, identified using the CT1 line, contains an 842 bp insert, the sequence of which is disclosed in SEQ ID NO:129. It contains partial ORFs for CT557 (Lpda) and CT558 (LipA).

Clone E3-F11-32, identified using the CT1 line, contains an 813 bp insert, the sequence of which is disclosed in SEQ ID NO:130. It contains a partial ORF for pmpD (CT812).

Clone E2-F8-5, identified using the CT12 line, contains a 1947 bp insert, the sequence of which is disclosed in SEQ ID NO:131. It contains a complete ORF for the 15 kDa ORF (CT442) and a partial ORF for the 60 kDa ORF (CT443).

Clone E2-G4-39, identified using the CT12 line, contains a 1278 bp insert, the sequence of which is disclosed in SEQ ID NO:132. It contains the partial ORF of the 60kDa ORF (CT443).

Clone E9-D1-16, identified using the CT10 line, contains a 916 bp insert, the sequence of which is disclosed in SEQ ID NO:133. It contains the partial ORF for the pmpH (CT872).

Clone E3-F3-6, identified using the CT1 line, contains a 751 bp insert, the sequence of which is disclosed in SEQ ID NO:134. It contains the partial ORFs, all on the complementary strand, for genes accB (CT123), L13 ribosomal (CT125), and S9 ribosomal (CT126).

Clone E2-D4-70, identified using the CT12 line, contains a 410 bp insert, the sequence of which is disclosed in SEQ ID NO:135. It contains the partial ORF for the pmpC gene (CT414).

Clone E5-A1-79, identified using the CT1 line, contains a 2719 bp insert, the sequence of which is disclosed in SEQ ID NO:136. It contains a partial ORF for ydhO (CT127), a complete ORF for S9 ribosomal gene (CT126 on the complementary strand), a complete ORF for the L13 ribosomal gene (CT125 on the complementary strand) and a partial ORF for accC (CT124 on the complementary strand).



Clone E1-F7-16, identified using the lines CT12, CT3, and CT11, contains a 2354 bp insert, the sequence of which is disclosed in SEQ ID NO:137. It contains a partial ORF of the *ftsH* gene (CT841) and the entire ORF for the *pnp* gene (CT842) on the complementary strand.

5 Clone E1-D8-62, identified using the CT12 line, contains an 898 bp insert, the sequence of which is disclosed in SEQ ID NO:138. It contains partial ORFs for the *ftsH* gene (CT841) and for the *pnp* gene (CT842).

#### EXAMPLE 4

##### EXPRESSION OF CHLAMYDIA TRACHOMATIS RECOMBINANT PROTEINS

10 Several *Chlamydia trachomatis* serovar E specific genes were cloned into pET17b. This plasmid incorporates a 6X histidine tag at the N-terminal to allow for expression and purification of recombinant protein.

15 Two full-length recombinant proteins, CT622 and CT875, were expressed in *E. coli*. Both of these genes were identified using CtLGVII expression screening, but the serovar E homologues were expressed. The primers used to amplify these genes were based on serovar D sequences. The genes were amplified using serovar E genomic DNA as the template. Once amplified, the fragments were cloned in pET-17b with a N-terminal  
20 6X-His Tag. After transforming the recombinant plasmid in XL-I blue cells, the DNA was prepared and the clones fully sequenced. The DNA was then transformed into the expression host BL21-pLysS cells (Novagen) for production of the recombinant proteins. The proteins were induced with IPTG and purified on Ni-NTA agarose using standard methods. The DNA sequences for CTE622 and CTE875 are disclosed in SEQ ID NO:28  
25 and 27 respectively, and their amino acid sequences are disclosed in SEQ ID NO: 140 and 139, respectively

Five additional *Chlamydia trachomatis* genes were cloned. The *Chlamydia trachomatis* specific protein CT694, the protein CT695, and the L1 ribosomal protein, the DNA sequences of which are disclosed in SEQ ID NO:119, 120 and 121 respectively. The  
30 protein sequences of these 6X-histidine recombinant proteins are disclosed in SEQ ID NO: 122 (CT694), 123 (CT695), and 124 (L1 ribosomal protein). The genes CT875 and

CT622, from serovar E were also cloned using pET17b as 6X-His fusion proteins. These recombinant proteins were expressed and purified and their amino acid sequences disclosed in SEQ ID NO:139 and 140, respectively.

5

### EXAMPLE 5

#### RECOMBINANT CHLAMYDIAL ANTIGENS RECOGNIZED BY T CELL LINES

Patient T cell lines were generated from the following donors: CT1, CT2, CT3, CT4, CT5, CT6, CT7, CT8, CT9, CT10, CT11, CT12, CT13, CT14, CT15, and CT16. A  
10 summary of their details is included in Table II.

<b>Table II: <i>C. trachomatis</i> patients</b>						
<b>Patients</b>	<b>Gender</b>	<b>Age</b>	<b>Clinical Manifestation</b>	<b>Serovar</b>	<b>IgG titer</b>	<b>Multiple Infections</b>
CT1	M	27	NGU	LCR	Negative	No
CT2	M	24	NGU	D	Negative	E
CT3	M	43	Asymptomatic Shed Eb Dx was HPV	J	Ct 1:512 Cp 1:1024 Cps 1:256	No
CT4	F	25	Asymptomatic Shed Eb	J	Ct 1:1024	Y
CT5	F	27	BV	LCR	Ct 1:256 Cp 1:256	F/F
CT6	M	26	Perinial rash Discharge, dysuria	G	Cp 1:1024	N
CT7	F	29	BV Genital ulcer	E	Ct 1:512 Cp	N

					1:1024	
CT8	F	24	Not Known	LCR	Not tested	NA
CT9	M	24	asymptomatic	LCR	Ct 1:128 Cp 1:128	N
CT10	F	20	Mild itch vulvar	negative	negative	12/1/98
CT11	F	21	BV Abnormal pap smear	J	Ct 1:512	F/F/J/E/E PID 6/96
CT12	M	20	asymptomatic	LCR	Cp 1:512	N
CT13	F	18	BV, gonorrhea, Ct vaginal discharge, dysuria	G	Ct 1:1024	N
CT14	M	24	NGU	LCR	Ct 1:256 Cp 1:256	N
CT15	F	21	Muco-purulent cervicitis Vaginal discharge	culture	Ct 1:256 Ct IgM 1:320 Cp 1:64	N
CT16	M	26	Asymptomatic/ contact	LCR	NA	N
CL8	M	38	No clinical history of disease	negative	negative	No

NGU=Non-Gonococcal Urethritis; BV=Bacterial Vaginosis; CT=Chlamydia trachomatis; Cp=Chlamydia pneumoniae; Eb=Chlamydia elementary bodies; HPV=human papilloma virus; Dx=diagnosis; PID=pelvic inflammatory disease; LCR=Ligase change reaction.

5

PBMC were collected from a second series of donors and T cell lines have been

generated from a sub-set of these. A summary of the details for three such T cell lines is listed in the table below.

Table III: Normal Donors				
Donor	Gender	Age	CT IgG Titer	CP IgG Titer
CHH011	F	49	1:64	1:16
CHH037	F	22	0	0
CHH042	F	25	0	1:16

- 5 Donor CHH011 is a healthy 49 year old female donor sero-negative for *C. trachomatis*. PBMC produced higher quantities of IFN-gamma in response to *C. trachomatis* elementary bodies as compared to *C. pneumoniae* elementary bodies, indicating a *C. trachomatis*-specific response. Donor CHH037 is a 22 year old healthy female donor sero-negative for *C. trachomatis*. PBMC produced higher quantities of IFN-
- 10 gamma in response to *C. trachomatis* elementary bodies as compared to *C. pneumoniae* elementary bodies, indicating a *C. trachomatis*-specific response. CHH042 is a 25 year old healthy female donor with an IgG titer of 1:16 to *C. pneumoniae*. PBMC produced higher quantities of IFN-gamma in response to *C. trachomatis* elementary bodies as compared to *C. pneumoniae* elementary bodies, indicating a *C. trachomatis*-specific response.
- 15 Recombinant proteins for several *Chlamydia trachomatis* genes were generated as described above. Sequences for MOMP were derived from serovar F. The genes CT875, CT622, pmp-B-2, pmpA, and CT529 were derived from serovar E and sequences for the genes gro-EL, Swib, pmpD, pmpG, TSA, CT610, pmpC, pmpE, S13, lpdA, pmpI, and pmpH-C were derived from LII.
- 20 Several of the patient and donor lines described above were tested against the recombinant Chlamydia proteins. Table IV summarizes the results of the T cell responses to the recombinant Chlamydia proteins.

**Table IV: Recombinant Chlamydia Antigens Recognized By T Cell Lines**

Antigen	Sero- var	#of hits	CL8 L2	CT10 E	CT1 E	CT3 E	CT4 L2	CT5 E	CT 11 E	CT 12 E	CT 13 E	CHH- 011 E	CHH- 037 E
gro-EL (CT110)	L2	10	-	+	+	+	+	+	+	+	+	+	+
MompF (CT681)	F	10	-	+	+	+	+	+	+	+	+	+	+
CT875	E	8	-	+	+	-	+	+	+	+	+	-	+
SWIB (CT460)	L2	8	+	+	-	+	-	+	-	+	+	+	+
pmpD (CT812)	L2	5	-	+	+	+	+	-	-	+	+	-	-
pmpG (CT871)	L2	6	-	+	+	-	+	+	nt	-	+	+	-
TSA (CT603)	L2	6	-	-	+	+	+	+	-	-	+	-	+
CT622	E	3	-	-	+	-	+	-	-	-	+	-	-
CT610	L2	3	-	+	-	+	-	-	-	+	-	-	-
pmpB-2 (CT413)	E	3	-	-	+	+	+	-	-	-	-	-	-
pmpC (CT414)	L2	4	-	-	-	+	-	+	-	+	-	-	+
pmpE (CT869)	L2	3	-	+	+	-	-	-	-	-	+	-	-
S13 (CT509)	L2	2	+	-	-	-	+	-	-	-	-	-	-
lpdA (CT557)	L2	3	-	-	+	+	-	-	-	-	-	+	-
pmpI (CT874)	L2	2	-	-	+	-	-	-	-	-	-	+	-
pmpH-C (CT872)	L2	1	-	-	-	-	-	-	-	+	-	-	-
pmpA	E	0	-	-	-	-	-	-	-	-	-	-	-

(CT412)													
CT529	E	0	-	-	-	-	-	-	-	-	-	-	-

### EXAMPLE 6

#### CD4 T CELL EXPRESSION CLONING FOR IDENTIFICATION OF T CELL STIMULATING ANTIGENS FROM CHLAMYDIA TRACHOMATIS SEROVAR E

5

The T cell line CHH037 was generated from a 22 year-old healthy female sero-negative for Chlamydia. This line was used to screen the Chlamydia trachomatis serovar E library (essentially as described in Example 1). Using this T cell line, we describe the identification of 7 clones.

10

Clone E8-D1-46 contains a 1754 bp insert, the sequence of which is disclosed in SEQ ID NO:143. It contains an almost complete ORF for the pepA gene (CT045) on the complementary strand, lacking a few amino acids towards the carboxy terminal end.

15

Clone E10-A11-10 contains a 3035 bp insert, the sequence of which is disclosed in SEQ ID NO:144. It contains partial ORFs for the yscU gene (CT091) and the truB gene (CT094) on the complementary strand and complete ORFs for the ychF gene (CT092) on the complementary strand and for the ribF gene (CT093) on the complementary strand.

20

Clone E8-B12-80 contains a 1353 bp insert, the sequence of which is disclosed in SEQ ID NO:145. It contains a short fragment of the SET domain protein gene (CT737) in frame with  $\beta$ -gal, a complete ORF in the complementary strand for the ybcL gene (CT736) as well as a partial ORF for the dag\_2 gene (CT735) on the complementary strand.

25

Clone E2-A8-70 contains a 1627 bp insert, the sequence of which is disclosed in SEQ ID NO:146. It contains a partial ORF for the mutS gene (CT792), a complete ORF for the ybcL gene (CT736) on the complementary strand, in addition to a partial ORF for the dag\_2 gene (CT735).

Clone E10-C1-47 contains a 1262 bp insert, the sequence of which is disclosed in SEQ ID NO:147. It contains a partial ORF for yaeI (CT461) on the

complementary strand, a complete ORF for SWIB (CT460) on the complementary strand and a partial ORF for prfB (CT459) on the top strand.

Clone E8-G7-86 contains a 1596 bp insert, the sequence of which is disclosed in SEQ ID NO:148. It contains a partial ORF for the mesJ (CT840) that is in frame with  $\beta$ -gal and a second partial ORF for the ftsH gene (CT841).

Clone E3-E6-84 contains a 2624 bp insert, the sequence of which is disclosed in SEQ ID NO:149. It contains a partial ORF for the pmpC gene (CT414) as well as a partial ORF on the complementary strand for the hypothetical gene CT611.

A second line, CHH042, which was generated from a healthy 25 year old female donor, seronegative for Chlamydia, was also screened against the Chlamydia trachomatis serovar E library. This screen led to the identification of 2 clones, E8-C12-38 and E1-D12-36.

Clone E8-C12-38 contains a 788 bp insert, the sequence of which is disclosed in SEQ ID NO:141. It contains partial ORFs for sfhB (CT658) and for the hypothetical gene, CT659.

Clone E1-D12-36 contains a 976 bp insert, the sequence of which is disclosed in SEQ ID NO:142. It contains a partial ORF for merB (CT709) in frame with  $\beta$ -gal, as well as a second partial ORF for the pckA gene (CT710).

## EXAMPLE 7

### CD4 T CELL EXPRESSION CLONING FOR IDENTIFICATION OF T CELL STIMULATING ANTIGENS FROM CHLAMYDIA TRACHOMATIS SEROVAR E

The T cell line CHH037 was generated from a 22 year-old healthy female sero-negative for Chlamydia. This line was used to screen the Chlamydia trachomatis serovar E library (essentially as described in Example 1). Using this T cell line, we describe the identification of clone E8-G7-54. This clone was found to contain a 3957 bp, the sequence of which is disclosed in SEQ ID NO: 157. It contains a partial ORF for the ftsH gene (CT841), which is in frame with  $\beta$ -gal. Clone E8-G7-54 also contains 2 partial ORFs on the complementary strand, for pGP7-D and pGP5-D, as well as a complete ORF

for pGP6-D, all three of which were from plasmid sequence.

A second T cell line, CHH042, which was generated from a healthy 25 year old female donor, seronegative for Chlamydia, was also screened against the Chlamydia trachomatis serovar E library. Using this T cell line, we describe the identification of 7 clones.

Clone E2-C3-27 contains a 1157 bp insert, the sequence of which is disclosed in SEQ ID NO:156. This clone contains complete ORFs for the genes rS3 (CT522) and rL22 (CT523) as well as partial oRFs for the genes rL16 (CT521) and rS19 (CT524).

Clone E10-F12-42 contains a 1909 bp insert, the sequence of which is disclosed in SEQ ID NO:155. It contains partial ORFs for the genes rS3 (CT522) and rL23 (CT526) as well as complete ORFs for the genes rL22 (CT523) rS19 (CT524) and rL2 (CT525).

Clone E10-F12-58 contains a 2275 bp insert, the sequence of which is disclosed in SEQ ID NO:154. It contains partial ORFs for the genes mhpA (CT148), rL16 (CT521), and rL2 (CT525) as well as complete ORFs for the genes rS3 (CT522), rL22 (CT523), and rS19 (CT524).

Clone E10-A8-16 contains a 3141 bp insert, the sequence of which is disclosed in SEQ ID NO:153. It contains partial ORFs for the genes rS3 (CT522) and rL3 (CT528) as well as complete ORFs for the genes rL22 (CT523), rS19 (CT524), rL2 (CT525), rL23 (CT526), and rL4 (CT527).

Clone E4-G8-49 contains a 1326 bp insert, the sequence of which is disclosed in SEQ ID NO:152. It contains partial ORFs for the genes pckA (CT710) and mreB (CT709), as well as a partial ORF for the pGP2-D from the plasmid.

Clone E9-E6-4 contains a 725 bp insert, the sequence of which is disclosed in SEQ ID NO:151. It contains a complete ORF for the hypothetical protein CT659 and a partial ORF for gyrA-2 (CT660).

Clone E2-A11-49 contains a 2052 bp insert, the sequence of which is disclosed in SEQ ID NO:150. It contains partial ORFs for the HAD superfamily (CT103) and the hypothetical protein, CT105, as well as a complete ORF for fabI (CT104).



Although the present invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, changes and  
5 modifications can be carried out without departing from the scope of the invention which is intended to be limited only by the scope of the appended claims.